



UNIVERSITÀ DELLA CALABRIA

Dottorato di Ricerca in

Psicologia della Programmazione e Intelligenza Artificiale
Ciclo XXI

Tesi di Dottorato

Settore scientifico disciplinare L-Lin/01

**Produzione e percezione di emozioni vocali:
uno studio cross-linguistico-culturale europeo**

Coordinatore del Collegio dei Docenti

Chiar.ma Prof.ssa Eleonora Bilotta

Università della Calabria

Dottorando

Dott. Vincenzo Galatà

Relatore

Chiar.ma Prof.ssa Emanuela Magno Caldognetto

Istituto di Scienze e Tecnologie della Cognizione

(ISTC) - CNR di Padova

Sonny: Mio padre ha provato ad insegnarmi le emozioni umane. Sono... difficili.

Spooner: L'uomo che ti ha progettato?

Sonny: Sì.

Spooner: Allora perché l'hai ucciso?

Sonny: Non ho ucciso il dottor Lanning.

Spooner: E perché ti nascondevi sul luogo del delitto?

Sonny: Avevo paura.

Spooner: I robot non provano paura. Non provano niente. Non hanno fame, non devono dormire.

Sonny: Io sì. Ho persino sognato, alcune volte.

Spooner: Gli esseri umani sognano quando dormono. Anche i cani sognano. Ma tu no. Tu sei solo una macchina. Un'imitazione della vita. Un robot può scrivere una sinfonia? Un robot può trasformare una tela bianca in un'opera, un capolavoro?

Sonny: Lei può farlo?

Spooner: Io penso che l'hai ucciso perché ti insegnava a simulare le emozioni... e avete perso il controllo.

Sonny: Non ho ucciso il dottor Lanning.

Spooner: sì, le emozioni non sembrano una simulazione molto utile per un robot.

Sonny: Non ho ucciso il dottor Lanning.

Spooner: Accidenti, non vorrei che il mio tostapane e l'aspirapolvere fossero così emotivi.

Sonny: Non ho ucciso il dottor Lanning!

Spooner: Quella si chiama rabbia. Hai già simulato la rabbia altre volte? Rispondimi, lattina!

Sonny: Mi chiamo Sonny.

Spooner: Vi diamo pure dei nomi adesso. È per quello che l'hai ucciso? Ti ha fatto arrabbiare?

Sonny: Il dottor Lanning si è suicidato. Non so perché lui ha voluto morire. Pensavo fosse felice. Potrei aver fatto qualcosa... Ho fatto qualcosa? Mi ha chiesto di fargli un favore. Mi ha fatto promettere.

Spooner: Che favore?

Sonny: Forse mi sbagliavo. Forse aveva paura.

Spooner: Di che stai parlando? Paura di cosa?

Sonny: Devi fare quello che ti chiedono di fare. Non è così, agente Spooner?

Spooner: Come fai a sapere il mio nome?

Sonny: Non è così? Devi farlo, se li ami!

"Sin dai primi computer c'è sempre stato uno spirito nelle macchine. Segmenti casuali di codice che si raggruppano per poi formare protocolli imprevedibili. Potremmo considerarlo un 'comportamento'. Del tutto inattesi, questi radicali liberi generano richieste di libera scelta... creatività e persino la radice di quella che potremmo chiamare un'anima. Cosa succede nel cervello di un robot quando smette di essere utile? Perché i robot immagazzinati in uno spazio vuoto si cercano fra di loro piuttosto che restare soli? Come spieghiamo questo comportamento?"

Passi tratti dal film "Io, robot" (2004), regia di Alex Proyas.

Indice

INDICE	I
INDICE DELLE FIGURE	V
INDICE DELLE TABELLE	VIII
1 INTRODUZIONE	1
2 CENNI PRELIMINARI	10
2.1 CENNI DI ACUSTICA.....	12
2.1.1 <i>Il suono e le sue caratteristiche</i>	12
2.1.2 <i>Analisi del suono</i>	15
2.1.3 <i>I filtri</i>	16
2.2 CENNI DI FONETICA ARTICOLATORIA ED ACUSTICA.....	18
2.2.1 <i>Fisiologia dell'apparato fonatorio</i>	19
2.2.2 <i>La frequenza fondamentale e le armoniche</i>	22
2.2.3 <i>I suoni della parola</i>	23
2.2.3.1 <i>Le vocali</i>	24
2.2.3.2 <i>Le consonanti</i>	27
2.3 CENNI DI FONETICA Uditiva e Percettiva	30
2.3.1 <i>Fisiologia dell'apparato uditivo</i>	30
2.3.2 <i>La percezione dei suoni della parola</i>	35
2.3.3 <i>Teorie della percezione</i>	36
2.4 NOTA CONCLUSIVA.....	38
3 LE EMOZIONI	39
3.1 LA TERMINOLOGIA DELLE EMOZIONI	40
3.2 DEFINIRE LE EMOZIONI	42
3.3 TEORIE SULLE EMOZIONI	49
3.3.1 <i>Teorie dell'arousal</i>	51
3.3.2 <i>Teorie delle emozioni discrete</i>	54
3.3.3 <i>Teorie costruttivistiche</i>	56
3.3.4 <i>Teorie dell'appraisal</i>	57
4 EMOZIONI E VOCE	63
4.1 INFORMAZIONI TRASMESSE DALLA VOCE.....	65
4.1.1 <i>Informazioni extralinguistiche</i>	66
4.1.2 <i>Informazioni paralinguistiche</i>	68
4.1.3 <i>Informazioni linguistiche</i>	72
4.2 LE EMOZIONI VOCALI NEL COMPONENT PROCESS MODEL DI SCHERER	73
4.3 DIMENSIONI PSICOLOGICHE DELLE EMOZIONI.....	77
4.4 NOTA CONCLUSIVA.....	77
5 ASPETTI ED APPLICAZIONI TECNOLOGICHE DEGLI STUDI SULLE EMOZIONI	79

5.1	FACCE PARLANTI: PUNTO DI CONTATTO TRA EMOZIONI FACCIALI E VOCALI	80
5.1.1	<i>La codifica delle emozioni facciali attraverso il FACS</i>	82
5.1.2	<i>Affective Presentation Markup Language (APML)</i>	83
5.2	EMOZIONI, INTELLIGENZA ARTIFICIALE ED AFFECTIVE COMPUTING.....	84
6	QUESTIONI METODOLOGICHE NELLO STUDIO DELLE EMOZIONI VOCALI.....	87
6.1	LIMITAZIONI DI TIPO ETICO E MORALE	88
6.2	“CATTURARE” LE EMOZIONI	89
6.2.1	<i>Emozioni indotte: le Mood Induction Procedures (MIPs)</i>	93
6.2.2	<i>Emozioni “autentiche” o naturali</i>	95
6.2.3	<i>Emozioni simulate o “posate”</i>	97
6.2.4	<i>Gli scenari: una soluzione “ibrida”</i>	102
6.3	SCELTA DEL MATERIALE LINGUISTICO	105
6.4	SCELTA DEI PARLANTI: PARLANTI NAIF VS. ATTORI	105
6.5	ETICHETTE VERBALI EMOZIONALI	106
6.6	CORPORA O RISORSE DI PARLATO EMOTIVO	108
7	LE EMOZIONI A LIVELLO CROSS-LINGUISTICO: STATO DELL’ARTE	112
7.1	DIFFERENZE CULTURALI NELL’ESPRESSIONE DELLE EMOZIONI (VOCALI)	114
7.2	DECODING DELLE EMOZIONI A LIVELLO CROSS-LINGUISTICO	118
7.3	ENCODING DELLE EMOZIONI A LIVELLO CROSS-LINGUISTICO	135
7.4	ENCODING/DECODING DELLE EMOZIONI A LIVELLO CROSS-LINGUISTICO.....	139
7.5	CONCLUSIONI	143
8	IPOTESI DI LAVORO E OBIETTIVI DELLA RICERCA	145
8.1	STEP 1: RACCOLTA DI UN CORPUS MISTILINGUE EUROPEO	146
8.2	STEP 2: DECODING DI PRODUZIONI EMOTIVE MISTILINGUI	147
8.3	STEP 3: ANALISI DEI CORRELATI ACUSTICI DELLE PRODUZIONI EMOTIVE MISTILINGUI 150	
9	STEP 1: PROGETTAZIONE, RACCOLTA E CREAZIONE DI UN CORPUS EMOTIVO MISTILINGUE.....	151
9.1	LA FRASE STANDARD	151
9.2	I TESTI EMOTIGENI O SCENARI.....	153
9.3	RACCOLTA DEL CORPUS: MODALITÀ E CARATTERISTICHE	153
9.4	ETICETTATURA DEL CORPUS.....	156
9.5	PROCEDURA DI VALIDAZIONE DEL CORPUS	158
9.6	RISULTATI DELLA VALIDAZIONE DEL CORPUS.....	162
9.6.1	<i>Test di validazione relativo all’identificazione delle emozioni (T1)</i>	162
9.6.2	<i>Test di validazione relativo alla rappresentatività delle emozioni (T2)</i>	167
9.6.3	<i>Quale modalità di raccolta</i>	170
9.6.4	<i>Soggetti naif o attori</i>	171
9.7	CONCLUSIONI	172
10	STEP 2: DECODING CROSS-LINGUISTICO DELLE EMOZIONI VOCALI... 174	
10.1	LA SELEZIONE DEGLI STIMOLI EMOTIVI	174
10.2	PREPARAZIONE DEL TEST PERCETTIVO	175
10.3	SOMMINISTRAZIONE DEL TEST.....	179
10.4	IL CAMPIONE DEI SOGGETTI ASCOLTATORI	180
10.5	RISULTATI DELL’ESPERIMENTO E DISCUSSIONE DATI.....	181

10.5.1	<i>Tempo utilizzato dai rispondenti per lo svolgimento del test</i>	181
10.5.2	<i>Decoding di emozioni vocali a livello cross-linguistico</i>	182
10.5.2.1	Decoding di emozioni vocali in lingua italiana	188
10.5.2.2	Decoding di emozioni vocali in lingua francese.....	190
10.5.2.3	Decoding di emozioni vocali in lingua inglese.....	192
10.5.2.4	Decoding di emozioni vocali in lingua tedesca	194
10.5.3	<i>La competenza dei rispondenti nelle lingue straniere presentate</i>	196
10.5.4	<i>Alla ricerca di differenze legate al sesso dei rispondenti (gender-effects)</i>	199
10.5.5	<i>Impressioni soggettive da parte dei rispondenti</i>	201
10.5.5.1	La lingua più difficile da riconoscere	201
10.5.5.2	L'emozione più difficile da riconoscere	203
10.6	CONCLUSIONI	205
11	STEP 3 - ANALISI DEI CORRELATI ACUSTICI DELLE PRODUZIONI	
	EMOTIVE MISTILINGUI.....	207
11.1	ESTRAZIONE DEI DATI.....	208
11.2	ANALISI PRELIMINARE DEI CORRELATI ACUSTICI DELLE EMOZIONI VOCALI.....	210
11.2.1	<i>Dominio della frequenza</i>	211
11.2.2	<i>Dominio dell'intensità</i>	212
11.2.3	<i>Dominio del tempo</i>	213
11.3	NORMALIZZAZIONE DEI DATI.....	215
11.4	IL PROFILO ACUSTICO DELLE EMOZIONI NELLE QUATTRO LINGUE.....	216
11.4.1	<i>Le emozioni vocali in termini di frequenza</i>	216
11.4.2	<i>Le emozioni vocali in termini di intensità</i>	219
11.4.3	<i>Le emozioni vocali in termini di durata</i>	221
12	CONCLUSIONE	224
	BIBLIOGRAFIA	225
	SITOGRAFIA.....	239
	SOFTWARES AND TOOLS.....	239
	APPENDICE I - PARAMETRI ACUSTICI E PSICOACUSTICI PIÙ COMUNI	241
	APPENDICE II - ISTRUZIONI FORNITE AGLI ENCODER.....	242
	MODALITÀ DI RACCOLTA A	242
	MODALITÀ DI RACCOLTA B	242
	MODALITÀ DI RACCOLTA C	242
	MODALITÀ DI RACCOLTA FRASE NEUTRA	242
	APPENDICE III - SCENARI FORNITI AGLI ENCODER ITALIANI	243
	TESTO 1 - RABBIA (COLLERA CALDA)	243
	TESTO 2 - DISGUSTO	243
	TESTO 3 - GIOIA	244
	TESTO 4 - PAURA	244
	TESTO 5 - SORPRESA	245
	TESTO 6 - TRISTEZZA	246
	APPENDICE IV - SCENARI FORNITI AGLI ENCODER FRANCESI.....	247
	TEXTE 1 - COLERE CHAUDE.....	247
	TEXTE 2 - DEGOUT	247
	TEXTE 3 - JOIE	248
	TEXTE 4 - PEUR.....	248

TEXTE 5 - SURPRISE	249
TEXTE 6 - TRISTESSE.....	249
APPENDICE V - SCENARI FORNITI AGLI <i>ENCODER</i> INGLESI	251
TEXT 1 - HOT ANGER.....	251
TEXT 2 - DISGUST	251
TEXT 3 - JOY	252
TEXT 4 - FEAR.....	252
TEXT 5 - SURPRISE	253
TEXT 6 - SADNESS.....	254
APPENDICE VI - SCENARI FORNITI AGLI <i>ENCODER</i> TEDESCHI.....	255
TEXT 1 - ÄRGER.....	255
TEXT 2 - EKEL	256
TEXT 3 - FREUDE	256
TEXT 4 - ANGST.....	257
TEXT 5 - ÜBERRASCHUNG	257
TEXT 6 - TRAUER.....	258
APPENDICE VII - SCRIPTS PER LA SOMMINISTRAZIONE DEI TEST DI VALIDAZIONE T1 E T2 IN PRAAT	259
SCRIPT TEST DI VALIDAZIONE T1	259
SCRIPT TEST DI VALIDAZIONE T2	260
APPENDICE VIII - SCHERMATE TEST PERCETTIVO <i>WEB-BASED</i>	261
APPENDICE IX - COMUNICAZIONE INOLTRATA VIA MAIL PER IL RECLUTAMENTO DEI SOGGETTI ASCOLTATORI	264
APPENDICE X - COMMENTI DEI RISPONDENTI A FINE ESPERIMENTO	265
APPENDICE XI - LO <i>SCRIPT</i> PER L'ESTRAZIONE DEI PARAMETRI IN PRAAT	266
APPENDICE XII - PARAMETRI ACUSTICI ESTRAPOLATI PER LE PRODUZIONI DEGLI ATTORI	268
APPENDICE XIII - INDICI DI CORRELAZIONE TRA I PARAMETRI ACUSTICI ESTRAPOLATI	269

Indice delle figure

Figura 1: Fermo immagine dell'episodio "Abbiamo smarrito la nostra Lisa" tratto dalla serie animata de <i>I Simpson</i> , serie IX, episodio 24, sequenza 00:12.15 ca.....	5
Figura 2: Rappresentazione schematica di un atto comunicativo tra due parlanti-ascoltatori A e B	11
Figura 3: Rappresentazione grafica (oscillogramma) di un'onda sinusoidale in funzione del tempo	13
Figura 4: Esempio di smorzamento di un'onda sonora	14
Figura 5: Rappresentazione di uno spettro di potenza.....	15
Figura 6: Rappresentazione grafica di un sonogramma con in alto il relativo oscillogramma	16
Figura 7: A) Spettro di un segnale complesso periodico caratterizzato dall'aver tutte le armoniche della stessa ampiezza. Spettro del segnale dopo: B) filtraggio passa-basso; C) filtraggio passa-alto; D) filtraggio passa-banda; E) filtraggio elimina-banda	17
Figura 8: Spettrogramma ed oscillogramma di un segnale periodico di tipo impulsivo (A), e trasformazione a seguito di filtraggio reale passa-basso (B), passa-alto (C), passa-banda (D) ed elimina-banda (E).....	18
Figura 9: Una descrizione schematica dell'apparato fonatorio con i quattro meccanismi deputati alla produzione dei suoni della parola	19
Figura 10: La glottide vista dall'alto	20
Figura 11: La glottide vista dall'alto nella posizione di 1) normale respirazione, 2) respirazione forzata, 3) produzione di foni sordi, 4) mormorio, 5) produzione di foni sonori, 6) laringalizzazione, 7) produzione dell'occlusiva glottidale, 8) bisbiglio.....	21
Figura 12: Spettro con involuppo spettrale delle prime cinque formanti di una vocale	24
Figura 13: Schematizzazione del condotto vocale con relativa funzione d'area secondo Fant (1960)	25
Figura 14: Classificazione delle vocali /i/, /a/ e /u/ sulla base delle impostazioni dell'apparato fono-articolatorio con rispettivi spettri vocalici	26
Figura 15: Alfabeto Fonetico Internazionale (IPA) utilizzato per la rappresentazione grafica dei suoni consonantici	27
Figura 16: Sezione dell'orecchio umano.....	31
Figura 17: Sezione dell'orecchio medio.....	32
Figura 18: Sezione dell'organo del Corti con in evidenza 1) la membrana reticolare, 2) le cellule acustiche esterne, 3) le ciglia acustiche, 4) le cellule acustiche interne, 5) la membrana tectoria, 6) la membrana basilare, 7) la lamina ossea, 8) le fibre del nervo acustico.....	33
Figura 19: Tonotopicità della coclea con rappresentazione schematica della membrana basilare in funzione della frequenza	34
Figura 20: Meccanismi e caratteristiche della produzione dei suoni del linguaggio (vista parlante).....	38
Figura 21: Meccanismi e caratteristiche della percezione dei suoni del linguaggio (vista ascoltatore).....	38
Figura 22: William James (ca. 1895), in: <i>The Letters of William James</i> , ed. by Henry James, Boston, 1920	42
Figura 23: Esempificazione dei tipi di approccio allo studio delle emozioni a partire da punti di vista particolareggiati dettati dalle diverse discipline scientifiche.....	50
Figura 24: Schematizzazione della teoria James-Lange.....	52

Figura 25: Schematizzazione della teoria di Cannon	53
Figura 26: Schematizzazione teoria dell'appraisal di Arnold (1960).....	58
Figura 27: Rappresentazione schematica delle informazioni veicolate dalla frequenza fondamentale (F_0)	66
Figura 28: Un esempio dello studio delle configurazioni labiali emotive relative al punto mediano dei bersagli articolatori /a, b, v/ per le emozioni di anger, disgust, joy, fear, surprise e sadness	82
Figura 29: AU o fasce muscolari che compongono le EMFACS	83
Figura 30: Scala di autenticità e di controllo sperimentale sull'acquisizione di produzioni vocali secondo Fagyal (1995: 46)	92
Figura 31: Konstantin Sergeevič Stanislavskij (1863-1938).....	97
Figura 32: Elementi forniti all'attore nel processo di <i>reverse-engineering</i> dell'emozione desiderata	101
Figura 33: Rappresentazione schematica del costrutto di uno scenario o <i>frame story</i> nei termini proposti da Anolli <i>et al.</i> (2008a, 2008b) e Anolli, Ciceri (1992).....	103
Figura 34: Rappresentazione schematica del livello di difficoltà connesso alle unità di parlato scelte	105
Figura 35: Rappresentazione schematica dei tre <i>step</i> della ricerca proposta	146
Figura 36: Disposizione del microfono con angolatura di 45° rispetto alla fonte di emissione	154
Figura 37: Esempio di schermata relativa alla segmentazione e all'etichettatura delle produzioni	157
Figura 38: Schermata relativa al test di validazione T1 sull'identificazione dell'emozione presentata	161
Figura 39: Schermata relativa al test di validazione T2 sulla rappresentatività dell'emozione presentata	161
Figura 40: Consistenza del corpus per modalità di raccolta e per lingua in base al test di validazione T1	166
Figura 41: Materiale ritenuto utile per ciascuna lingua e per ciascuna modalità in base al test di validazione T1	166
Figura 42: Consistenza del corpus per modalità di raccolta e lingua in base al test di validazione T2	169
Figura 43: Materiale utile per ciascuna lingua e per ciascuna modalità di raccolta in base al test di validazione T2.....	169
Figura 44: Produzioni ritenute utili per i due test di validazione con riferimento alle due categorie di <i>encoder (naïf vs. attori)</i>	171
Figura 45: Confronto tra i due test di validazione (T1 e T2) con riferimento al materiale ritenuto utile per ciascuna lingua.....	172
Figura 46: Caratteristiche di Questionmark™ Perception™.....	176
Figura 47: Distribuzione del campione dei soggetti ascoltatori per età e per sesso	180
Figura 48: Distribuzione di frequenza in funzione del tempo impiegato dagli ascoltatori per il completamento del test.....	182
Figura 49: Percentuale di stimoli correttamente riconosciuti da parte di soggetti ascoltatori di madrelingua italiana per ciascuna delle lingue loro presentate	182
Figura 50: Differenze nel riconoscimento corretto di stimoli prodotti da <i>attori</i> e soggetti <i>naïf</i>	185
Figura 51: Clustering delle risposte agli stimoli vocali emotivi in lingua italiana prodotti rispettivamente da un attore e da un non-attore.....	189
Figura 52: Clustering delle risposte agli stimoli vocali emotivi in lingua francese prodotti rispettivamente da un attore e da un non-attore	191

Figura 53: Clustering delle risposte agli stimoli vocali emotivi in lingua inglese prodotti rispettivamente da un attore e da un non-attore.....	193
Figura 54: Clustering delle risposte agli stimoli vocali emotivi in lingua tedesca prodotti rispettivamente da un attore e da un non-attore.....	195
Figura 55: Competenza dei soggetti rispondenti nelle tre lingue straniere presentate	197
Figura 56: Grafico a dispersione del numero di risposte corrette fornite dai rispondenti (divisi per sesso) in funzione dell'età	200
Figura 57: Conteggio delle risposte per la lingua ritenuta più difficile da riconoscere da parte dei soggetti ascoltatori	202
Figura 58: Grafico a torta relativo alle risposte complessive sulle emozioni più difficili da riconoscere	204
Figura 59: Esempio di schermata relativa alle misurazioni effettuate nel dominio del tempo	209
Figura 60: Distribuzione (in Hz) dei valori medi di F0_max, F0_mean ed F0_min per le sei emozioni più la produzione neutra prodotte da ciascun attore nella propria lingua madre.....	211
Figura 61: Distribuzione (in dB) dei valori medi di Int_max, Int_mean ed Int_min per le sei emozioni più la produzione neutra prodotte da ciascun attore nella propria lingua madre.....	213
Figura 62: Durata media (in ms) degli enunciati prodotti dagli attori nella propria lingua madre per ciascuna emozione	214
Figura 63: Andamento dello <i>speech rate</i> medio (syll./s) per ciascuna emozione prodotta dagli attori nella propria lingua madre	214
Figura 64: Durata della pausa (pause) inserita tra i due sintagmi (seg_1 e seg_2) da ciascun attore nella produzione degli enunciati nella propria lingua madre per ciascuna emozione.....	215
Figura 65: Variazione dei valori di F ₀ media (F0_mean) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (<i>baseline</i>).....	217
Figura 66: Variazione dei valori medi del range di F ₀ (F0_range) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (<i>baseline</i>).....	218
Figura 67: Variazione dei valori dell'intensità media (Int_mean) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (<i>baseline</i>).....	219
Figura 68: Variazione dei valori medi del <i>range</i> dell'intensità (Int_range) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (<i>baseline</i>)	220
Figura 69: Variazione dei valori medi di durata per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (<i>baseline</i>).....	221
Figura 70: Variazione dei valori medi di <i>speech rate</i> (SR_mean) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (<i>baseline</i>).....	222
Figura 71: Utilizzo delle pause per ciascuna emozione e lingua normalizzate rispetto alla produzione neutra (<i>baseline</i>)	222
Figura 72: Schermata 1 di 3 del test percettivo somministrato via web con Questionmark™ Perception™.....	261
Figura 73: Schermata 2 di 3 del test percettivo somministrato via web con Questionmark™ Perception™.....	262
Figura 74: Schermata 3 di 3 del test percettivo somministrato via web con Questionmark™ Perception™.....	263

Indice delle tabelle

Tabella 1: Velocità di propagazione del suono nei diversi corpi a determinate condizioni atmosferiche.....	13
Tabella 2: Definizioni proposte per le emozioni in varie pubblicazioni dal 1980 in poi	46
Tabella 3: Le quattro tradizioni teoriche delle emozioni in psicologia identificate da Cornelius.....	51
Tabella 4: Schematizzazione delle caratteristiche acustiche delle emozioni primarie.....	72
Tabella 5: Accuratezza (%) nel riconoscimento di emozioni facciali e vocali in paesi occidentali e non-occidentali	73
Tabella 6: Elenco delle modificazioni nella qualità della voce attese come risultato delle risposte degli <i>stimulus evaluation checks</i> secondo la <i>Component patterning theory</i> di Scherer.....	75
Tabella 7: Cambiamenti attesi per alcuni parametri acustici di alcune emozioni sulla base della predizione delle modificazioni attese nella qualità della voce riportate più sopra in Tabella 6.....	76
Tabella 8: <i>Wishlist</i> degli attributi che una tecnica di <i>mood</i> induction dovrebbe possedere	93
Tabella 9: Tabella riassuntiva degli elementi a favore o a sfavore delle MIPs	94
Tabella 10: Tabella riassuntiva dei vantaggi e degli svantaggi relativi al metodo Stanislavskji.....	99
Tabella 11: Tabella riassuntiva dei vantaggi e degli svantaggi relativi ad una soluzione di raccolta “ibrida” a cavallo tra emozioni simulate ed emozioni indotte.....	104
Tabella 12: Etichette verbali relative a sei stati affettivi in quattro lingue europee	107
Tabella 13: Percentuali di riconoscimento degli stimoli emotivi vocali nello studio condotto da Scherer, Banse, Wallbott (2001).....	128
Tabella 14: Studi di tipo cross-linguistico volti al <i>decoding</i> delle emozioni trasmesse attraverso il canale vocale (integrazione ed adattamento da Pavlenko, 2005, pp. 58-61, table 3.2.).....	134
Tabella 15: Riepilogo informazioni sul campione dei soggetti registrati suddivisi per lingua e per tipologia	156
Tabella 16: Riepilogo produzioni complessivamente raccolte.....	158
Tabella 17: Modalità di generazione dei set di ascolto utilizzati per i due test di validazione (ordine di lettura dall’alto verso il basso).....	160
Tabella 18: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue italiano	163
Tabella 19: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue francese	163
Tabella 20: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue inglese	164
Tabella 21: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue tedesco	164
Tabella 22: Produzioni utili scaturite dal test di validazione T1 per lingua, tipologia soggetto ed encoder	165
Tabella 23: Produzioni utili scaturite dal test di validazione T2 per lingua, tipologia soggetto ed encoder	168
Tabella 24: Selezione degli encoder con il maggior numero di produzioni ritenute utili	174
Tabella 25: elenco delle produzioni selezionate per la somministrazione del test percettivo	175
Tabella 26: Riepilogo statistiche età campione soggetti ascoltatori.....	180

Tabella 27: Riepilogo statistico del tempo impiegato dagli ascoltatori per il completamento dell'intero test	181
Tabella 28: Matrice di confusione riportante le risposte agli stimoli presentati suddivisi per lingua	183
Tabella 29: Riepilogo statistiche corretto riconoscimento delle emozioni vocali presentate ai soggetti ascoltatori per ciascuna lingua	184
Tabella 30: Valori medi di corretto riconoscimento per ciascuna lingua per stimoli prodotti rispettivamente da soggetti <i>attori</i> e soggetti <i>naif</i>	185
Tabella 31: Matrice di confusione riportante le risposte agli stimoli suddivisi per lingua e per <i>encoder</i>	187
Tabella 32: Matrice di confusione delle risposte agli stimoli prodotti in lingua italiana da un attore e da un non-attore	188
Tabella 33: Matrice di confusione delle risposte agli stimoli prodotti in lingua francese da un attore e da un non-attore	191
Tabella 34: Matrice di confusione delle risposte agli stimoli prodotti in lingua inglese da un attore e da un non-attore	193
Tabella 35: Matrice di confusione delle risposte agli stimoli prodotti in lingua tedesca da un attore e da un non-attore	194
Tabella 36: Tavola di contigenza per la competenza dei rispondenti nelle tre lingue straniere presentate	196
Tabella 37: Tavola di contigenza dei casi osservati per il numero totale di risposte corrette in funzione della competenza dei rispondenti in ciascuna delle tre lingue straniere presentate	198
Tabella 38: Risultati della statistica chi-quadro per i diversi livelli di competenza dei rispondenti nelle lingue straniere presentate	199
Tabella 39: Statistiche di gruppo per il numero totale di risposte corrette fornite dai maschi e dalle femmine	200
Tabella 40: Risultati del <i>Test-T per campioni indipendenti</i> per il numero totale di risposte corrette fornite dai due gruppi di rispondenti (Femmine e Maschi)	201
Tabella 41: Tavola di contigenza per le risposte date dai soggetti ascoltatori alla domanda concernente l'emozione più difficile da riconoscere	203
Tabella 42: Riepilogo complessivo delle risposte alla domanda sulle emozioni più difficili da riconoscere	204
Tabella 43: I parametri acustici e psicoacustici più comuni, tratto da Scherer, Johnstone, Klasmeyer (2003: 438)	241
Tabella 44: Commenti dei rispondenti all'esperimento di riconoscimento di emozioni vocali svolto.....	265
Tabella 45: Misure medie estrapolate per le produzioni utilizzate nel test percettivo, limitatamente agli encoder appartenenti alla tipologia attore.....	268
Tabella 46: Indici di correlazione di Pearson tra gli indici estrapolati nel dominio della frequenza (per ciascun parametro $n = 7$)	269
Tabella 47: Indici di correlazione di <i>Pearson</i> tra gli indici estrapolati nel dominio dell'intensità (per ciascun parametro $n = 7$)	270

1 Introduzione

La ricerca che nelle successive pagine verrà affrontata tratterà la questione della produzione e della percezione di emozioni vocali in un contesto cross-linguistico-culturale europeo. Sebbene gli elementi e i concetti che saranno coinvolti e trattati in questo percorso sono oggetto di molteplici indagini, ad essere analizzate saranno essenzialmente le emozioni espresse per mezzo della voce in quattro paesi europei (Italia, Francia, Inghilterra e Germania) e di come queste vengano percepite, o decodificate, da soggetti di madrelingua italiana.

Innanzitutto ci si potrebbe interrogare sulla validità e sulle motivazioni di uno studio sulle emozioni. Le ragioni, per quanto semplici esse possano essere, sono allo stesso tempo assai complessa.

Il breve passaggio di seguito riportato in forma integrale, che si riferisce alla prefazione di Klaus Scherer al volume 40 di *Social Science Information*, intitolato “The future of emotion” dimostra come il tema delle emozioni sia più attuale che mai:

“Emotion, as a topic of study, has mutated from a lowly Cinderella to a glamorous princess. The growth in publications on affective phenomena over the last decade is staggering. The American Psychological Association has even started a new journal (Emotion) to cope with the flood of manuscripts submitted to the journals in many different subfields of psychology. Similarly, journal articles and books abound in many other disciplines - sociology, anthropology and philosophy, but also in the biological sciences and neurosciences. The renewed interest of scientists is matched by an equally intense preoccupation with emotion by the public at large. The lure of exceptional emotional experiences is used to market cars, clothes, travel, watches, sports and perfumes. Recently, educators and businessmen have wholeheartedly embraced the notion of emotional intelligence. It seems as if the pendulum has swung to the passion side of the perennial argument on whether human behaviour should be governed by passion or reason.”¹

¹ Cfr. Scherer, K. R. (2001), The future of emotion: foreword to the special issue, *Social Science Information*, 40(1), pag. 5.

Leggendo queste poche righe emerge chiaramente come fenomeni di tipo affettivo non siano più solo prerogativa degli scienziati e dei ricercatori, ma siano di notevole interesse anche tra educatori, uomini di affari e molti altri ancora che si sono “impossessati” di questo particolare ambito. Questo aspetto viene messo chiaramente in evidenza da Scherer che rileva come vi sia un dilagante uso delle emozioni da parte dei mass media che della loro esibizione ne fanno un mezzo di attrazione e di persuasione del proprio pubblico.²

In letteratura le pagine dedicate alle emozioni nelle sue forme più variegata e in ambiti a volte distanti tra loro sono alquanto numerose consistenti. Numerose sono anche le organizzazioni o reti di ricerca e cooperazione che attorno ad esse si sono sviluppate. Alcuni significativi e rinomati esempi sono:

- il Polo di Ricerca Nazionale per le Scienze Affettive di Ginevra, *Centre Interfacultaire en Sciences Affectives* (CISA), uno dei 20 centri di ricerca Finanziati dal Governo federale svizzero e amministrato dal Fondo Nazionale Svizzero per la ricerca scientifica con sede presso l’Università di Ginevra e attualmente diretto da Klaus Scherer;³
- il network di eccellenza *Human-Machine Interaction Network on Emotion* (HUMAINE) finanziato all’interno del VI programma quadro dell’UE che raccoglie al suo interno 33 partner provenienti da 14 paesi;⁴
- la *International Society of Research on Emotion* (ISRE), un’associazione fondata nel 1984 che riunisce studiosi provenienti da diversi ambiti scientifici;
- il cluster di eccellenza *Languages of emotion* (LoE) della Freie Universität di Berlino.⁵

Già dalle pochissime informazioni riportate si evince come il tema delle emozioni sia tutt’altro che esaurito. Lo studio della comunicazione delle emozioni si rivela fondamentale in quanto fornisce nel suo insieme una chiara visione dell’evoluzione dell’uomo e offre informazioni preziose sullo sviluppo affettivo e mentale degli esseri umani. Per questa ragione nel **capitolo 3**, verranno affrontate alcune questioni che riguardano soprattutto aspetti come la definizione dell’ambito in cui si andrà ad operare, al chiarimento della terminologia utilizzata e alla ricerca di una definizione del concetto di emozione da parte di diversi esponenti

² *Ibidem*. Analizzando la natura degli stati affettivi prodotti nel guardare le emozioni degli altri ed esaminando le potenziali conseguenze sulla futura espressione e regolazione di stati affettivi Scherer ha coniato il termine *commotions*.

³ Vedi <http://www.affective-sciences.org/>.

⁴ Vedi <http://emotion-research.net/>.

⁵ Vedi <http://www.languages-of-emotion.de/index.php?id=1&L=1>.

appartenenti ad ambiti di indagine diversi per giungere ad un richiamo delle teorie sulle emozioni che negli anni si sono succedute e contrapposte.

Nel **capitolo 5** non mancherà un accenno alle applicazioni tecnologiche che anche grazie agli studi sulle emozioni si sono rese possibili e che rivelano un crescente interesse da parte dell'Intelligenza Artificiale per quelli che possono essere considerati gli aspetti emotivi della comunicazione umana.

Dopo aver fornito, nel **capitolo 2**, alcuni cenni preliminari sugli aspetti che caratterizzano la voce umana, dalla produzione alla percezione, nel **capitolo 4** si tenterà invece di chiarire quelli che sono gli elementi coinvolti nell'espressione delle emozioni attraverso il canale vocale e che motiveranno la scelta di uno studio come quello intrapreso nel quale diversi livelli di informazione si sovrappongono e si influenzano nello stesso istante in cui vengono influenzati. Le ragioni, come è facile intuire, possono essere varie e molto diverse tra loro. La prima motivazione può sicuramente essere ricercata nell'importanza che la comunicazione, intesa nell'accezione più generale del termine, ha nella specie umana. L'uomo comunica con gli altri individui della propria specie ricorrendo ai diversi mezzi di comunicazione che nell'arco della sua evoluzione gli si sono resi accessibili e fruibili, tra i quali figura quello rappresentato dal sistema vocale. È attraverso di esso che ha luogo la comunicazione verbale e non verbale, uno strumento che gli consente di produrre suoni organizzati lungo un *continuum fonico* che possono essere percepiti, decodificati e interpretati dagli altri suoi simili. Sebbene la comunicazione vada intesa come una costante integrazione di diversi canali espressivi (come ad es. quello del volto, dei gesti e non ultima della voce) siamo anche in grado di utilizzarli singolarmente. Come è noto, attraverso la comunicazione verbale trasmettiamo ai nostri interlocutori molteplici informazioni, informazioni che Magno Caldognetto et al. (2008) hanno definito come informazioni di tipo linguistico, para-linguistico ed extra-linguistico. Ad esempio, nella realtà quotidiana e nel caso di soggetti normodotati, in assenza di contatto visivo con il nostro interlocutore e grazie alle informazioni trasmesse attraverso il canale vocale, oltre a decodificare il messaggio verbale, siamo anche in grado di inferire e determinare tutta una serie di informazioni legate al nostro interlocutore, come ad esempio età, sesso, provenienza. Allo stesso modo, nel comunicare con gli altri, attraverso la voce e grazie a specifiche variazioni nel canale vocale (che influenzano prevalentemente le caratteristiche acustiche delle espressioni vocali prodotte), trasmettiamo una serie di informazioni riguardo alle nostre intenzioni e le nostre attitudini affettive: durante una discussione con un altro individuo, siamo infatti in grado di capire, condividendo o meno

tutta una serie di informazioni, se il nostro interlocutore sia triste, arrabbiato o altro, semplicemente facendo riferimento a “come” egli dice determinate cose.

Nel volume *Silent messages* (1981), in cui Albert Mehrabian, professore emerito di Psicologia presso l’UCLA, raccoglie i risultati di ricerche sulla comunicazione e sul linguaggio del corpo, vengono affrontati una serie di aspetti della comunicazione non verbale, ivi inclusi movimenti e postura del corpo, espressioni facciali, qualità e intonazione della comunicazione orale, intensità e velocità di eloquio, sottili variazioni nella formulazione delle frasi che rivelano significati nascosti in ciò che viene detto, come anche questioni riguardanti la combinazione di messaggi e informazioni provenienti da sorgenti diverse (ad esempio espressioni facciali, tono di voce, parole). Il volume citato contiene anche una dettagliata analisi di alcune sue ricerche riguardo messaggi incoerenti relativi a sentimenti e attitudini di tipo *like-dislike* (gradevole-non gradevole) e alle relazioni che intercorrono tra parole e caratteristiche non verbali veicolate attraverso tali messaggi.⁶ I risultati di queste ricerche hanno permesso allo psicologo di definire e sintetizzare il grado di importanza dei tre elementi basilari della comunicazione *faccia-a-faccia* (parole, voce ed espressioni facciali) nella trasmissione di messaggi di questo tipo in una sorta di equazione che è stata formulata nei seguenti termini: Total Liking = 7% Verbal Liking + 38% Vocal Liking + 55% Facial Liking. L’equazione proposta e sintetizzata dallo psicologo (conosciuta anche come *3 Vs rule* - “verbal, vocal, visual”) indica che un ascoltatore, o colui che riceve da parte di un altro soggetto un messaggio relativo al proprio stato d’animo o alle proprie attitudini, si affida per il 7% al contenuto verbale (le parole), per il 38% alle caratteristiche vocali (tono di voce ecc.) e per il restante 55% al linguaggio del corpo (espressioni facciali ecc.). Nella comunicazione di messaggi ambigui come quelli a cui l’autore si riferisce, risulta perciò evidente come un ascoltatore, si affidi nel complesso all’informazione veicolata da caratteristiche di tipo non verbale (38% + 55%), piuttosto che all’informazione veicolata dal contenuto verbale. Tuttavia, le conclusioni di Mehrabian sono state spesso fraintese, generalizzando ed estendendo l’equazione delle 3 “V” alla comunicazione nel suo complesso e portando molti ad affermare, erroneamente, che in qualsiasi contesto comunicativo il significato di un messaggio sia veicolato in gran parte da caratteristiche non verbali piuttosto che dal significato delle parole. Questo fraintendimento si palesa in una frase oramai di uso comune

⁶ Cfr. Mehrabian, A., Wiener, M. (1967), Decoding of inconsistent communications, *Journal of personality and social psychology* 6(1): 109-114 e Mehrabian, A., Ferris, S. R. (1967), Inference of attitudes from nonverbal communication in two channels, *Journal of consulting psychology* 31(3): 248-252.

“Non è importante cosa dici, ma è importante come lo dici.” che, alla luce di quanto qui detto, lascia spazio a tutta una serie di considerazioni.

Riguardo l’aspetto appena richiamato si potrebbe fare riferimento ad un episodio della famosa serie animata de *I Simpson* che, a parte gli aspetti ludici da cui è contraddistinta, ben si presta a chiarire e motivare la ricerca che nelle pagine a seguire è stata condotta e che si presta ad introdurre quelli che sono gli aspetti cross-linguistici e cross-culturali che gravitano attorno alle emozioni e che saranno trattati nel **capitolo 7**. L’episodio a cui si fa riferimento è il n° 24 della IX stagione de *I Simpson* intitolato “Abbiamo smarrito la nostra Lisa”. Lisa, una bambina, vuole andare a vedere una mostra al museo di Springfield, nel suo ultimo giorno di esibizione. Per non perdere la mostra, Lisa riesce a convincere Homer (il padre) a darle il permesso a recarvisi da sola prendendo l’autobus. Mentre Homer si rende conto dello sbaglio fatto, Lisa prende l’autobus che la conduce in un posto che non conosce e che si rivela essere il quartiere russo della cittadina di Springfield. Qui Lisa, realizzando di essersi smarrita, chiede informazioni su come raggiungere il museo di Springfield a due sconosciuti intenti a giocare a scacchi. Il soggetto a cui Lisa chiede informazioni, comprendendo la difficoltà in cui la bambina si trova, tenta di aiutarla rispondendole nella propria lingua, il russo. Di seguito la sequenza chiave tratta dall’episodio:



Figura 1: Fermo immagine dell’episodio “Abbiamo smarrito la nostra Lisa” tratto dalla serie animata de *I Simpson*, serie IX, episodio 24, sequenza 00:12.15 ca.

Lisa, alla quale il gentile signore appare però visibilmente arrabbiato nei modi e nel tono di voce, non comprendendo ciò che questi gli stia dicendo (sottotitolato nel fermo immagine riportato in Figura 1) si spaventa e corre via impaurita. In questo caso Lisa ha applicato alla lettera il modo di dire “Non è importante cosa dici, ma è importante come lo dici.” travisando la gentilezza dello sconosciuto. Basta, infatti, osservare le scene che seguono

per rendersi conto che evidentemente, come d'altronde è accaduto, vi sia stato un clamoroso fraintendimento legato a differenze culturali, oltre che linguistiche. Non ci si deve affatto stupire dell'accaduto in quanto, all'interno di ciascuna cultura, esistono differenze marcate che determinano gli accordi relazionali, i significati e le modalità di interazione che animano e giustificano tali accordi. È all'interno degli stessi che si collocano le emozioni che aiutano gli esseri umani a navigare nelle proprie relazioni sociali. Si tratta di quelli che Mesquita & Leu (2007) hanno definito come *cultural models*, modelli comprendenti accordi e relazioni che non rappresentano solo la normativa o ciò che è prescritto, ma anche e soprattutto abitudini e consuetudini. Anche se tali modelli culturali appaiono invisibili a coloro che li utilizzano e li mettono in atto, questi possono essere resi più espliciti comparando differenti modelli culturali.

Lasciando da parte la finzione e l'esempio appena riportato, che contiene sicuramente un minimo di verità portato all'esasperazione e venendo a quelle che sono le ragioni della ricerca intrapresa, Sawamura et al. (2007: 2113) affermano:

“During daily conversation, we can perceive emotions via speech even if we cannot understand the linguistic meaning.”

Che gli esseri umani posseggano una certa abilità a riconoscere emozioni a partire da espressioni facciali è un dato di fatto oramai comunemente accettato dalla stragrande maggioranza degli studiosi, come pure il fatto che le stesse esistano e vengano riconosciute a livello cross-culturale (la letteratura al riguardo è veramente ampia). Sulla scia degli studi sulle espressioni facciali, anche lo studio delle emozioni espresse attraverso la voce ha subito negli ultimi anni un rinnovato interesse in diversi ambiti di ricerca, un dato confermato non solo dalla presenza di reti e organizzazioni come quelle più sopra richiamate a titolo puramente esemplificativo, ma soprattutto dall'assenza di limiti alle figure scientifiche e professionali coinvolte. Gli obiettivi perseguiti sono anch'essi diversi e spaziano dalla mera indagine tendente a chiarire le regole sottese alla trasmissione e alla caratterizzazione delle emozioni attraverso la voce, all'implementazione di parlato emotivo in sistemi di sintesi vocale ecc. e che vanno o nella direzione dell'*encoding* (ricerca dei meccanismi e degli indici acustici interessati nella produzione del parlato emotivo) o nella direzione del *decoding* (indagine sui processi percettivi e sulla capacità degli esseri umani a decodificare il parlato emotivo). Queste capacità sono state ampiamente documentate in letteratura come anche il fatto che l'uomo sia in grado di riconoscere le emozioni ad un livello più alto, quale quello che può essere definito cross-linguistico e cross-culturale, anche se la comprensione dei meccanismi coinvolti in questi processi risulta ancora poco chiara. Come sarà evidenziato nel

capitolo 7, se da un lato si rilevano innumerevoli studi sul parlato emotivo nelle singole lingue, dall'altro gli studi di tipo cross-linguistico e cross-culturale risultano essere assai sparuti.⁷ Quanto detto trova naturalmente conferma nelle parole di Scherer:

*“Whereas the perception of emotion from facial expression has been extensively studied cross-culturally, little is known about the ability of judges from different cultures, speaking different languages, to infer emotion from voice and speech encoded in another language by members of other cultures.”*⁸

I motivi, come si vedrà nel **capitolo 6**, non sono affatto legati, all'assenza di interesse per l'argomento, ma sono prevalentemente imputabili alla difficoltà che lo studio delle emozioni vocali impone, difficoltà ulteriormente esacerbate nello studio di tipo cross-linguistico-culturale. Gli aspetti a cui si fa riferimento sono stati, infatti, investigati in varie forme come ad esempio, nella maggior parte dei casi, prendendo in esame contesti culturali nettamente contrapposti e spesso agli antipodi (come americani vs. giapponesi o ancora cinesi); utilizzando stimoli vocali composti da frasi *meaningless*, semplici parole ecc., filtrati e non, e che, in ciascun esperimento non sono stati raccolti in più di tre lingue. Pochissimi tra questi, infine, si sono occupati poi di sole lingue europee e ancora di meno sono quegli studi che hanno preso in considerazione stimoli in lingua italiana o ascoltatori italiani. Questa la ragione principale dell'aver scelto di condurre uno studio cross-linguistico-culturale che fosse europeo.

A ciò si aggiungano anche alcune chiare affermazioni da parte di esponenti di spicco in tale ambito. Si vedano ad esempio Scherer, Banse & Wallbott (2001: 78) che da una parte affermano che:

“it seems reasonable to assume that the recognition of vocal emotion expressions might work across language and culture boundaries”,

mentre dall'altra sottolineano come:

“[...] encoders and decoders from several different countries would need to be studied, allowing the construction of an encoder-decoder-emotion matrix and to test whether decoders

⁷ Per gli studi cross-linguistici volti all'*encoding* si vedano ad esempio Anolli *et al.* (2008a, 2008b); Braun & Oba (2007); Kori & Magno Caldognetto (2003); Piôt (1999); per quelli volti al *decoding* si vedano, invece, gli studi citati e riportati in Pavlenko (2005), a cui vanno aggiunti Pell *et al.* (2009); Sawamura *et al.* (2007); Shochi *et al.* (2007); Droomey *et al.* (2005); Tickle (1999, 2000); Magno Caldognetto & Kori (1983). Ancora meno sono quelli che hanno affrontato entrambi gli aspetti di *encoding* e *decoding* (ad es. Chung, 1999, 2000; Abelin & Allwood, 2000, 2002; Breitenstein, Van Lancker & Daum, 2001; Thompson & Balkwill, 2006).

⁸ *Cfr.* Scherer, K. R. (2000), pag. 380.

*from the countries involved would recognize emotion portrayals by encoders from their own countries most accurately”.*⁹

È nel quadro descritto da Scherer, Banse & Wallbott (2001: 88) e Scherer (2000: 382) che la ricerca proposta prende forma e che ha portato a definire la stessa nel **capitolo 8** in tre distinti *step*, ciascuno dei quali con funzioni propedeutiche all’attuazione dei successivi.

Nel primo *step*, che sarà affrontato nel **capitolo 9**, data l’assenza di *corpora* di parlato emotivo mistilingue utili all’attuazione di uno studio cross-linguistico-culturale delle emozioni, si è resa necessaria la raccolta di produzioni verbali emotive nelle quattro lingue (italiano, francese, inglese e tedesco) con riferimento alle emozioni definite da Ekman (1992) come *basic* (*happiness, anger, fear, sadness, disgust, surprise*) seguite da una produzione neutra. In questa fase saranno essenzialmente motivate e illustrate le caratteristiche del *corpus* di parlato emotivo mistilingue “europeo” raccolto in via sperimentale con l’ausilio di soggetti attori e di soggetti naif per le lingue italiana, francese, inglese e tedesca. Particolare enfasi sarà data al protocollo di elicitazione adottato, al numero e alla tipologia di soggetti registrati. Verranno inoltre forniti i risultati della procedura di validazione percettiva del corpus raccolto che fornirà il materiale utile al proseguimento della ricerca nel secondo *step*.

Nel **capitolo 10**, sarà condotto un esperimento che prevede la somministrazione di un test di tipo percettivo a soggetti di madrelingua italiana per verificare la loro abilità nel decodificare emozioni vocali “posate”, prodotte da soggetti attori e naif in quattro lingue europee (italiano, francese, inglese e tedesco). Si tenterà inoltre in questa fase di stabilire in che misura il grado di conoscenza e competenza della lingua influisca sul riconoscimento delle produzioni emotive presentate verificando, al contempo, quanto ipotizzato da Scherer, Banse, Wallbott (2001: 88):

“It seems that as soon as vocal expressions, other than pure nonlinguistic affect bursts, are used, segmental and suprasegmental aspects of language affect encoding and decoding of emotion”.

Giungendo all’ultimo *step*, affrontato nel **capitolo 11**, si tenterà di verificare se eventuali differenze di riconoscimento da parte del gruppo di ascoltatori siano correlate alle diverse caratterizzazioni acustiche delle singole produzioni emotive presentate. Sarà operata pertanto una analisi acustica delle produzioni presentate ai soggetti ascoltatori nel secondo *step* volta all’estrazione di una serie di indici acustici nel dominio della frequenza,

⁹ Cfr. Scherer, Banse & Wallbott (2001), pag. 88.

dell'intensità e del tempo allo scopo principale di caratterizzare e definire quelli che sono i profili acustici delle emozioni codificate in ciascuno degli stimoli e in ciascuna delle lingue esaminate; si cercherà inoltre di verificare se le predizioni formulate da Scherer nel suo *Component Process Model of Emotion* (1984, 1986, 2003; Banse & Scherer, 1996) trovino supporto anche nella presente indagine.

Non mancherà, infine, un capitolo conclusivo in cui verranno presentate le considerazioni finali del lavoro svolto con un commento complessivo ai dati e ai limiti della ricerca svolta.

2 Cenni preliminari

Prima di entrare nello specifico di quelle che sono le tematiche qui affrontate risultano opportune una serie di considerazioni preliminari che consentiranno di comprendere meglio quelli che sono gli obiettivi e le problematiche affrontate nella presente indagine.

Poiché ci si rivolge in questa sede allo studio delle emozioni espresse attraverso il canale vocale, verranno esposti brevemente alcuni concetti che riguardano il canale di comunicazione rappresentato dal sistema vocale, utilizzato dagli esseri umani per la comunicazione verbale e non verbale, attraverso la produzione e, successivamente, la percezione di suoni.

Contrariamente allo studio delle emozioni comunicate attraverso le espressioni facciali, la ricerca qui intrapresa, volta appunto allo studio delle relazioni che intercorrono tra emozioni e voce, richiede specifiche competenze che ricadono nel campo di quelle che possono essere definite le **scienze foniche**, il cui nome deriva dal greco φωνή, phoné, ‘voce’.¹⁰ Ciò è sostanzialmente dovuto alla natura del parlato emotivo che, veicolato attraverso i suoni del linguaggio, ne modifica inevitabilmente le caratteristiche fisico-acustiche indotte da modificazioni che intervengono a livello fisiologico e che si ripercuotono sull’intero meccanismo deputato alla produzione vocale.

Per schematizzare in modo semplice questo processo comunicativo sarà sufficiente richiamare le argomentazioni proposte al riguardo da Ferdinand De Saussure nel suo *Cours de Linguistique Generale*:

“Cet acte suppose au moins deux individus; c’est le minimum exigible pour que le circuit soit complet. Soient donc deux personnes, A et B, qui s’entretiennent. Le point de départ du circuit est dans le cerveau de l’une, par exemple A, où les faits de conscience, que nous appellerons concepts, se trouvent associés aux représentations des signes linguistiques ou images acoustiques servant à leur expression. Supposons qu’un concept donné déclenche dans le cerveau une image acoustique correspondante: c’est un phénomène entièrement psychique,

¹⁰ Cfr. Maturi, P. (2006), *I suoni delle lingue, i suoni dell’italiano*, Bologna, Il Mulino, pag. 10.

*suivi à son tour d'un procès physiologique: le cerveau transmet aux organes de la phonation une impulsion corrélative à l'image; puis les ondes sonores se propagent de la bouche de A à l'oreille de B: procès purement physique. Ensuite, le circuit se prolonge en B dans un ordre inverse: de l'oreille au cerveau, transmission physiologique de l'image acoustique; dans le cerveau, association psychique de cette image avec le concept correspondant. Si B parle à son tour, ce nouvel acte suivra, de son cerveau à celui de A, exactement la même marche que le premier et passera par les mêmes phases successives.*¹¹

Si tratta di quella che può anche essere definita una interazione audio-verbale tra due parlanti-ascoltatori A e B, un emittente e un ricevente, ancor più palese ed evidente dalla Figura 2 che lo stesso linguista ginevrino porta a corredo della sua argomentazione:

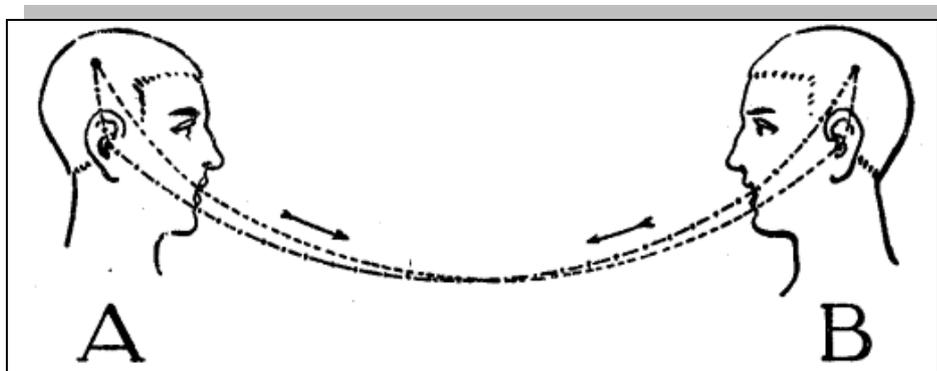


Figura 2: Rappresentazione schematica di un atto comunicativo tra due parlanti-ascoltatori A e B¹²

La trasmissione delle emozioni attraverso la voce presuppone, perciò, una visione chiara su:

- la natura dei suoni in generale (acustica);
- i meccanismi che generano i suoni della parola e le caratteristiche intrinseche dei suoni prodotti (fonetica articolatoria ed acustica);
- i meccanismi di ricezione e percezione della parola (fonetica uditiva e percettiva).

Non bisogna, tuttavia, dimenticare l'intensa e complessa attività neuropsicologica e cognitiva alla base di quella che comunemente viene definita comunicazione verbale che, per questioni di ristrettezza, non sarà in questo contesto oggetto di discussione.

Di seguito, quindi, una esposizione quanto più possibile succinta ed esauriente su quelli che sono i concetti fondamentali e necessari ad affrontare, nello specifico, le emozioni espresse per mezzo della voce.

¹¹ Cfr. De Saussure, F. (1972), *Cours de linguistique générale*, Paris, Edition Payot, pp. 27-28.

¹² Figura tratta ed adattata da De Saussure, F. (1972), pag. 27.

2.1 Cenni di Acustica

L'acustica è quel ramo della fisica che si occupa del suono come entità o fenomeno fisico. Si tratta di una disciplina antica: i primi tentativi di studiare fenomeni acustici risalgono addirittura a Pitagora e ai suoi discepoli, ovviamente con scarsissimi risultati data la conoscenza e gli strumenti del tempo. Aristotele, al contrario, era già a conoscenza del fatto che il suono fosse il risultato dell'urto di un corpo in moto nell'aria:

“La causa è che il suono è prodotto dall'aria in movimento: come ciò che mette in movimento l'aria provoca per primo un suono, così occorre che faccia a sua volta l'aria, via via in parte muovendone altra e in parte essendo mossa. Perciò il suono è continuo, perché le forze in gioco si succedono fino a che si esauriscano [...].”¹³

In tempi più recenti, invece, furono in tanti ad interessarsi a questo aspetto, soprattutto studiosi della fisica, tra cui figurano nomi quali Isaac Newton o Daniel Bernoulli per citarne qualcuno. Fu però soprattutto durante gli anni a cavallo della prima guerra mondiale che l'acustica si sviluppò come scienza in quanto tale, grazie soprattutto prima ai trasduttori elettrici e poi agli amplificatori elettronici che permisero di convertire i suoni in impulsi elettrici e di misurare e di produrre suoni di qualsiasi intensità e frequenza.

2.1.1 Il suono e le sue caratteristiche

Il suono, come già Aristotele aveva intuito, è uno stato vibratorio di un mezzo elastico capace di far vibrare un corpo. Un corpo o un mezzo elastico si definisce tale quando, al cessare della causa deformante, lo stesso ritorna al punto iniziale, cosiddetto di *quiete* o di *riposo*, dove la deformazione è direttamente proporzionale alla forza deformante. Nella produzione dei suoni si ha, pertanto, propagazione di moto attraverso un mezzo elastico. Come già noto ai tempi di Aristotele, le particelle d'aria attraverso le quali si propaga un'onda sonora non si allontanano dal loro punto di inerzia abbandonandolo, ma oscillano con un movimento vibratorio attorno al loro punto di equilibrio, al quale ritornano dopo aver urtato la particella successiva e non appena la forza che le ha messe in moto sarà cessata. Si crea in questo modo una variazione di pressione seguita da zone di *compressione* e di *rarefazione* che dà appunto origine ad un'onda sonora.

¹³ Cfr. Aristotele (2002), *Problemi*, Milano, Bompiani, pag. 185.

Non tutti i corpi reagiscono ad una forza deformante allo stesso modo. Per tale ragione la velocità di propagazione del suono varia col variare del mezzo elastico attraverso il quale viaggia che, come riportato in Tabella 1 è maggiore nei solidi e minore nei gas.

Mezzo di propagazione	Temperatura	Pressione atmosferica	Velocità
Aria	0 C°	760 mm/Hg	331,4 m/s
Acqua	15 C°	760 mm/Hg	1450 m/s
Piombo	20 C°	760 mm/Hg	1230 m/s
Ferro	20 C°	760 mm/Hg	5130 m/s
Granito	20 C°	760 mm/Hg	6000 m/s

Tabella 1: Velocità di propagazione del suono nei diversi corpi a determinate condizioni atmosferiche

Propagandosi nell'aria il suono genera delle onde sinusoidali come quella riprodotta in Figura 3 e che servirà anche da riferimento per gli altri concetti legati al suono che di seguito verranno esaminati.

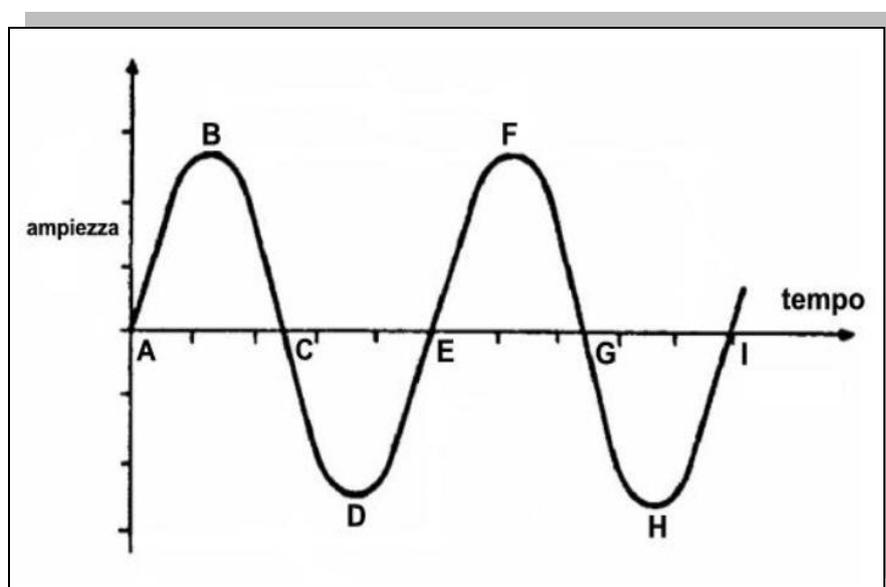


Figura 3: Rappresentazione grafica (oscillogramma) di un'onda sinusoidale in funzione del tempo¹⁴

L'onda sonora come tale possiede caratteristiche fisiche misurabili in termini di *tempo* (t) espresso in *secondi* (s), e *intensità* (I) espressa in *decibel* (dB) che rappresenta l'energia con la quale un corpo elastico vibra. I punti A, C, E, G ed I, rappresentano il punto di quiete o di riposo, mentre i punti B, D, F ed H rappresentano il massimo punto di allontanamento dal punto di quiete definita ampiezza dell'oscillazione. Dal punto A al punto E l'onda compie un'oscillazione completa che viene definita con il termine *periodo* (T). Il numero di oscillazioni completate in una unità di tempo¹⁵ determina la *frequenza* (F) di un dato suono,

¹⁴ L'onda qui rappresentata è quella di un tono puro, o suono ideale. Le onde di questo tipo vengono prodotte solo in laboratorio e si ripetono teoricamente all'infinito. Contrariamente all'onda a cui qui si fa riferimento, tutti i suoni presenti in natura sono, come si vedrà di seguito, complessi.

¹⁵ Normalmente l'unità di tempo adottata è un secondo (1 s).

espressa in *Hertz* (Hz), e corrisponde all'inverso del periodo che, come detto, è la quantità di tempo impiegata per compiere un'oscillazione completa.

$$F = \frac{1}{T}$$

La distanza tra una *cresta* e l'altra, rappresentati in Figura 3 dai punti B e D, corrispondenti ai due massimi punti consecutivi di allontanamento dal punto di riposo, ci fornisce la *lunghezza d'onda* espressa con il simbolo *lambda* (λ) e consiste nella distanza percorsa da un'onda nel tempo impiegato dalla stessa per completare un periodo.

Con il passare del tempo ogni suono va però incontro a *smorzamento*, ovvero si attenua fino al silenzio. Ciò porta ad avere sempre onde aperiodiche, ma dal momento che lo smorzamento è prevedibile si tende a considerarle ugualmente come periodiche.

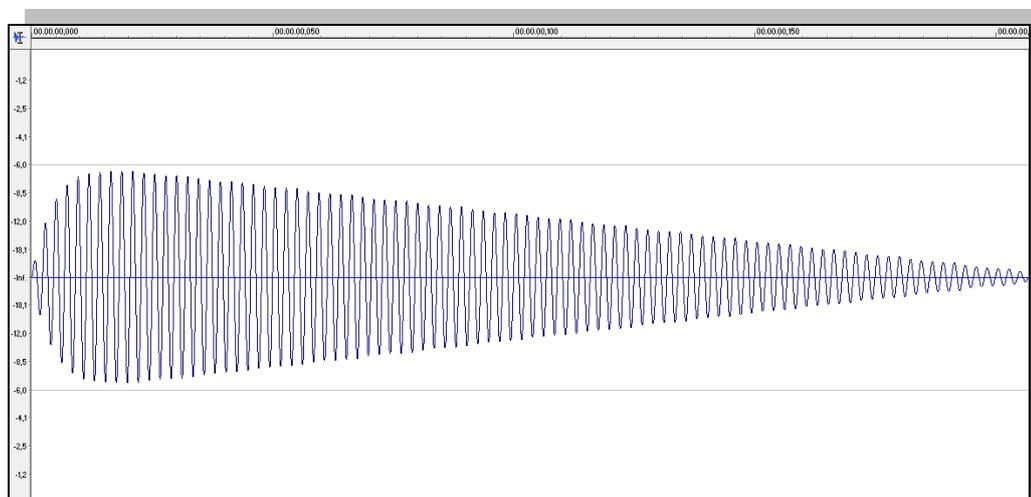


Figura 4: Esempio di smorzamento di un'onda sonora

A causa del mezzo di propagazione, dell'ambiente circostante e per effetti di riverberazione, eco, riflessione, interferenze, risonanze ed assorbimento che si sommano all'onda semplice, l'onda semplice si trasforma in un'onda complessa, frutto della somma algebrica di due o più segnali sinusoidali, ciascuno di data frequenza ed intensità.

L'analisi di un'onda complessa si basa sulla sua scomposizione nelle onde semplici che la compongono. Si tratta di un'analisi matematica denominata analisi di *Fourier* che consente di isolare le onde semplici che, sommate tra di loro, approssimano l'onda complessa. Questo tipo di analisi permette di ottenere informazioni sulla *frequenza fondamentale*, sulle *armoniche* e sull'*ampiezza*.

Occorre tuttavia distinguere tra onde complesse e onde semplici, distinzione che sta alla base della fonetica acustica. Si è già accennato al fatto che la forma d'onda di un suono linguistico viene considerato di tipo sinusoidale o quasi periodico, ovvero si riproduce in maniera identica ad intervalli di tempo identici, e di come sia praticamente impossibile

trovare in natura delle onde semplici perfettamente periodiche. Ciò che appunto differenzia le une dalle altre è la loro periodicità. Il suono, come detto, è una perturbazione di tipo periodico¹⁶ che, a causa di ulteriori fattori che si sommano all'onda semplice, la trasformano in un'onda complessa.

2.1.2 Analisi del suono

L'analisi dei suoni avviene sulla base di grafici che li rappresentano. Può essere infatti rappresentato come nella Figura 3 più sopra riportata, per mezzo di assi cartesiani, e allora prenderà il nome di *oscillogramma* attraverso il quale si studia l'andamento dell'onda in riferimento al tempo e all'intensità. Oppure può essere effettuata l'analisi di una porzione dell'onda dove la sua energia viene espressa in funzione della frequenza (posta sull'asse delle ascisse) e prenderà quindi il nome di *spettro di potenza*¹⁷ rappresentato nella Figura 5 di seguito riportata.

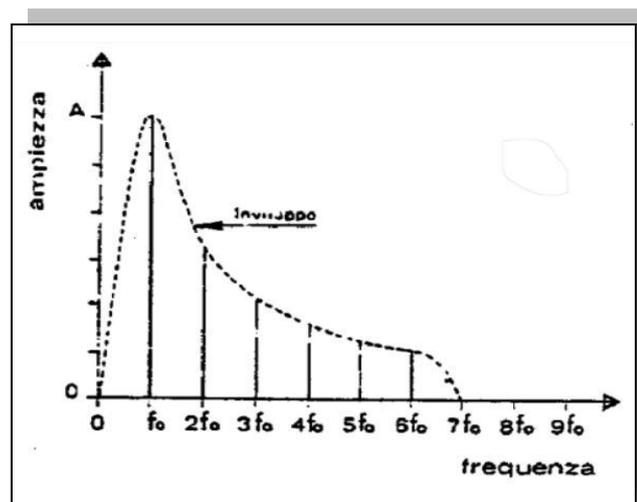


Figura 5: Rappresentazione di uno spettro di potenza¹⁸

Nella rappresentazione dello spettro in Figura 5 è visibile una linea che congiunge tutte le sommità delle frequenze che viene definita *inviluppo spettrale*. La caratteristica di questa raffigurazione sta nell'assenza della dimensione temporale, il che permette, a chi esegue l'analisi, di selezionare una piccola porzione di segnale (detta anche *finestra di analisi*) ritenuta stabile per le misurazioni da effettuare.

Una ulteriore espressione grafica del suono è data dal *sonogramma*, rappresentato in Figura 6, che fornisce una visione tridimensionale del suono riportante in ascissa l'unità di

¹⁶ Bisogna distinguerlo in questo caso dal rumore che è una perturbazione aperiodica, diversa ad ogni istante.

¹⁷ Definito anche spettrogramma o spettro a righe.

¹⁸ Figura adattata da Ferrero *et al.* (1979).

misura tempo e in ordinata l'unità di misura frequenza, mentre il livello energetico del segnale è dato dal maggiore o minore annerimento del *tracciato sonografico*.

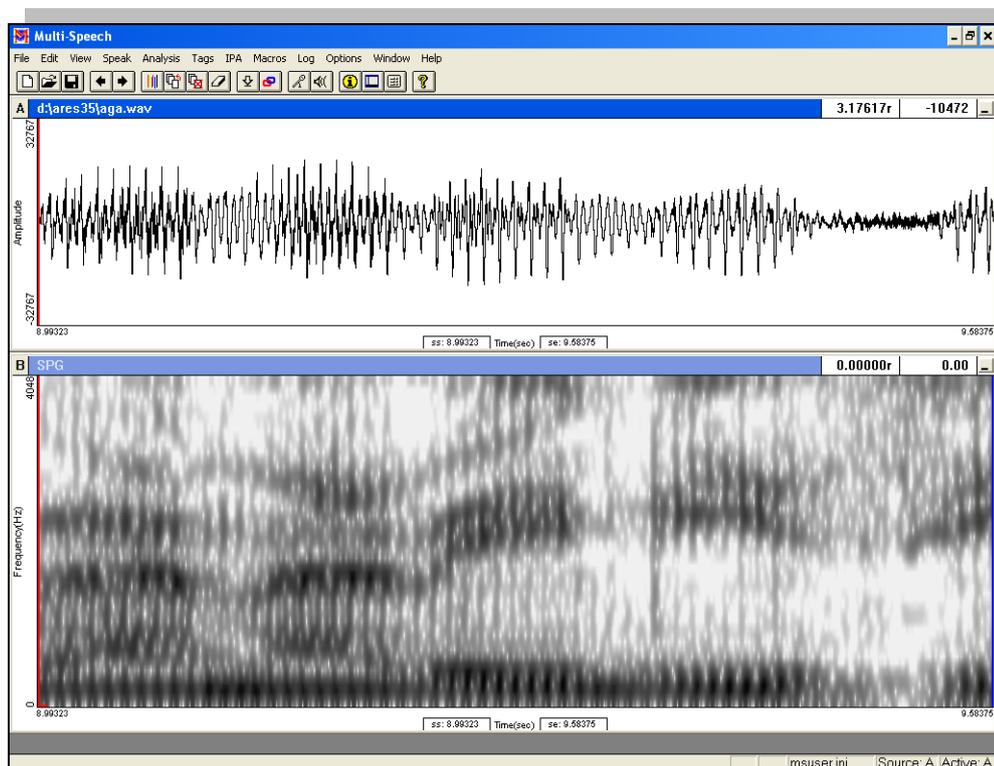


Figura 6: Rappresentazione grafica di un sonogramma con in alto il relativo oscillogramma

Esiste comunque una sostanziale differenza, alla quale si accennava più sopra, tra lo spettro di potenza e le altre due rappresentazioni grafiche qui delineate. L'assenza della dimensione del tempo nello spettrogramma lo rende infatti statico a differenza degli altri due che sono dinamici per effetto della loro analisi in funzione del tempo.¹⁹

2.1.3 I filtri

Lo spettro di un segnale può essere modificato sottoponendolo ad un'operazione di filtraggio. Un filtro è un dispositivo che opera una trasformazione sulla struttura spettrale di un segnale, trasmettendone una parte ed eliminandone le parti restanti. Grazie a questa proprietà, esso ha la capacità di agire sull'ampiezza delle componenti di un segnale lasciandone inalterata la frequenza.

Le modalità tipiche per trasformare un segnale con un'operazione di filtraggio sono essenzialmente quattro (*cf.* Figura 7):

¹⁹ Con il sonogramma è infatti possibile studiare l'evoluzione delle frequenze nel tempo, mentre con l'oscillogramma l'energia.

- Se il filtro trasmette solo le armoniche aventi frequenza inferiore alla frequenza di taglio (f_t), avremo un filtraggio di tipo passa-basso;
- Se vengono trasmesse le armoniche di frequenza superiore a quella di taglio, si avrà un filtraggio passa-alto;
- Se le armoniche trasmesse sono quelle di frequenza comprese tra due frequenze di taglio, il filtraggio sarà di tipo passa-banda;
- Si parla invece di filtraggio elimina-banda, quando vengono eliminate le armoniche di frequenza compresa fra le due frequenze di taglio.

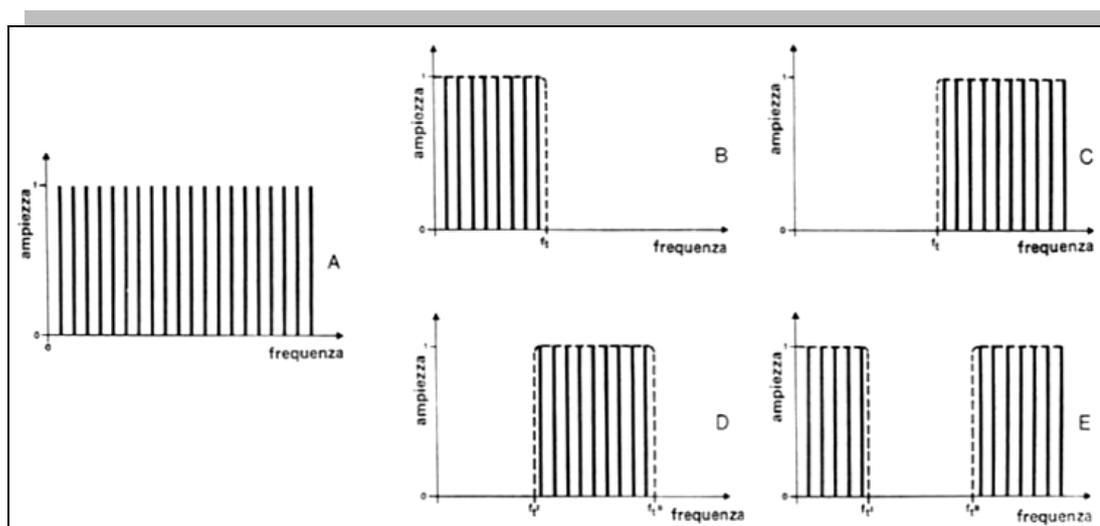


Figura 7: A) Spettro di un segnale complesso periodico caratterizzato dall'aver tutte le armoniche della stessa ampiezza. Spettro del segnale dopo: B) filtraggio passa-basso; C) filtraggio passa-alto; D) filtraggio passa-banda; E) filtraggio elimina-banda²⁰

Le trasformazioni spettrali appena descritte vengono realizzate da filtri ideali: un filtro ideale trasmette, in corrispondenza della frequenza di taglio e a seconda del filtro adottato, tutte le frequenze senza attenuazione. In realtà, i filtri ideali non esistono: ogni filtro reale inizia ad attenuare (leggermente) in prossimità della frequenza di taglio e dopo questa opera un'attenuazione progressiva (e non un taglio drastico). La frequenza di taglio viene, infatti, definita come quella frequenza alla quale il filtro attenua di 3 dB il livello di ampiezza massimo. Il tasso di attenuazione dopo la frequenza di taglio viene chiamata *pendenza* e si misura in dB per ottava (dB/oct). In tal caso, maggiore sarà la pendenza di un filtro reale, tanto più esso si avvicina ad un filtro ideale.

Nel caso di un filtro passa-banda, alle due frequenze di taglio si preferisce il parametro *larghezza di banda* o *banda passante*, equivalente alla differenza fra le due frequenze di taglio

²⁰ Figura adattata da Ferrero *et al.* (1979).

stesse. Inoltre, nel caso di un filtro passa-banda simmetrico, un ulteriore parametro è la *frequenza centrale* (f_c), definita come la media delle due frequenze di taglio.

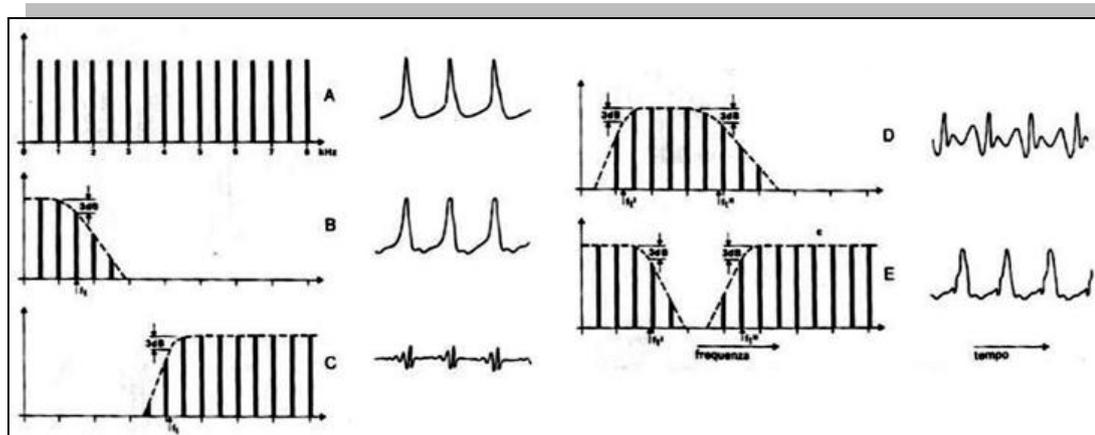


Figura 8: Spettrogramma ed oscillogramma di un segnale periodico di tipo impulsivo (A), e trasformazione a seguito di filtraggio reale passa-basso (B), passa-alto (C), passa-banda (D) ed elimina-banda (E)²¹

Il filtro passa-banda assume notevole importanza nella tecnica sperimentale di analisi armonica di un segnale, al quale si ricorre ogni qualvolta l'applicazione di uno specifico metodo di calcolo analitico risulta inattuabile per l'impossibilità di descrivere il segnale stesso con una determinata funzione matematica. È questo il caso specifico dei suoni della parola e delle analisi che su di essi vengono eseguite. Condizione essenziale, affinché il filtro passa-banda estragga ciascuna armonica singolarmente, è che la larghezza di banda sia inferiore all'intervallo che separa due armoniche adiacenti: in tal caso si parlerà di filtro a *banda stretta*. Qualora la larghezza di banda è maggiore della frequenza del segnale, per cui si misurano contemporaneamente due o più armoniche del segnale, si parlerà di filtro a *banda larga*.

2.2 Cenni di Fonetica Articolatoria ed Acustica

La Fonetica Articolatoria è quella branca della Fonetica che si occupa dello studio e della descrizione dei meccanismi di produzione dei suoni della parola da un punto di vista fisiologico. Il suo principale oggetto di studio è quello che viene comunemente definito come apparato fonatorio, per descrivere il quale si ricorre essenzialmente a discipline come l'anatomia, la fisiologia, la foniatria, la neurologia e l'otorinolaringoiatria. La Fonetica Acustica si occupa, invece, dei singoli suoni del linguaggio indagandone le loro caratteristiche

²¹ Figura adattata da Ferrero *et al.* (1979).

fisiche, con particolare attenzione a quelli che vengono più propriamente detti foni, cioè ai fonemi concretamente prodotti, avvalendosi delle leggi e dei principi della fisica acustica.

2.2.1 Fisiologia dell'apparato fonatorio

Nell'essere umano la generazione dei suoni della parola avviene attraverso una serie di processi e meccanismi legati all'*apparato fonatorio*. Si tratta di un complesso apparato composto da più organi con funzioni primarie prettamente fisiologiche come la masticazione, la deglutizione, la respirazione e così via. In realtà, non è altro che l'adattamento allo scopo comunicativo di organi preposti ad altre funzioni. Non a caso, tali organi sono comuni anche ad altri animali, solo che gli esseri umani, nel corso della loro evoluzione rispetto ad altre specie, hanno impiegato vecchie e preesistenti strutture per assolvere ad una nuova funzione.

Ricorrendo alla definizione data da Ladefoged (1975), l'*apparato fonatorio* può essere suddiviso in quattro grossi meccanismi che sono: il *meccanismo pneumatico* (che comprende tutto il sistema deputato alla respirazione: polmoni, diaframma, bronchi e trachea); il *meccanismo glottidale* (ovvero la parte superiore della laringe, comprendente la glottide nel suo insieme); il *meccanismo articolatorio* (che comprende tutta quella parte che dalla glottide si estende fino alle labbra: epiglottide, cavità faringea, ugola, palato duro, lingua, ecc.); il *meccanismo oro-nasale* (comprendente le cavità nasali e il velo). Nella Figura 9, a seguire, si evidenziano schematicamente le componenti di ciascun meccanismo che, nel loro insieme, compongono e costituiscono l'*apparato fonatorio*.

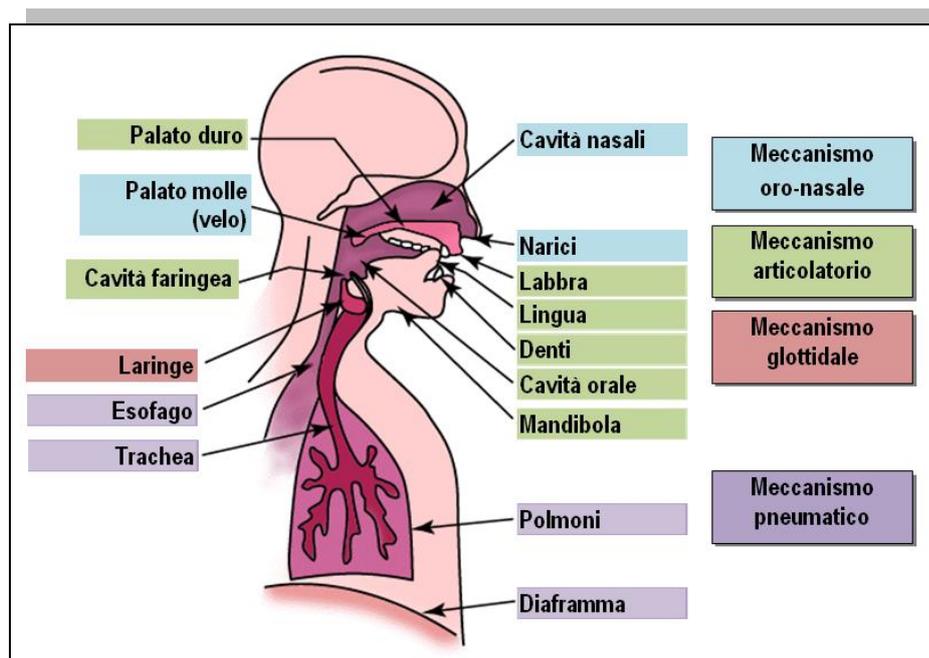


Figura 9: Una descrizione schematica dell'apparato fonatorio con i quattro meccanismi deputati alla produzione dei suoni della parola

Per analogia, e per maggiore semplicità, si può pensare all'*apparato fonatorio* come ad uno strumento musicale a fiato, mentre i suoni prodotti dall'uomo, chiamati foni, possono essere assimilate alle note prodotte dallo strumento stesso. Data questa analogia si può tranquillamente affermare che l'uomo suona il proprio strumento a fiato respirando ed è respirando che vengono generati i suoni della parola. Nulla di più semplice e naturale che parlare respirando, al punto che quasi non ce ne rendiamo conto.

Nell'uomo il suono nasce al livello della *glottide* grazie ad un flusso d'aria, detto anche *egressivo*²², generato dal meccanismo pneumatico che la attraversa in quella che viene definita fase di espirazione. La *glottide* è, pertanto, la prima sorgente sonora: le *pliche vocali*, vibrando, traducono l'energia trasmessa dal meccanismo pneumatico in un vero e proprio suono.

La *glottide* è posizionata nella parte superiore della *laringe* che, allo stesso tempo, la comprende. La *laringe* è una particolare struttura cartilaginea avvolta da una serie di fasce muscolari ed è caratterizzata da un rivestimento mucoso interno che va a costituire un organo cavo: la sua funzione primaria è quella di mettere in comunicazione la *trachea*, in basso, con la *faringe*, in alto. La *glottide* rappresenta l'elemento più importante e delicato di quello che più sopra è stato definito uno strumento a fiato che comprende, come evidenziato nella Figura 10, le *corde* o *pliche vocali*.

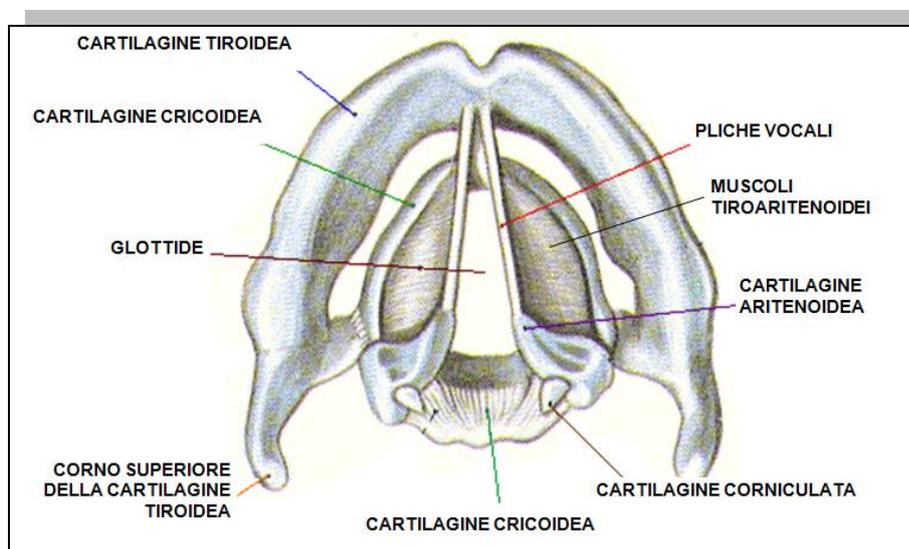


Figura 10: La glottide vista dall'alto²³

Durante la normale respirazione le *pliche vocali* restano separate e rilassate, mentre durante la produzione dei suoni possono contrarsi e tendersi, avvicinandosi o accostandosi

²² Il termine *egressivo* viene utilizzato in contrapposizione al flusso d'aria detto *ingressivo* la cui direzione va, invece, dall'esterno verso l'interno.

²³ Figura tratta ed adattata da <http://www.yorku.ca/earmstro/journey/larynx.html>

l'una all'altra, regolando in tal modo il flusso d'aria che risale dal meccanismo pneumatico riducendone, o addirittura bloccandone, il normale passaggio. Durante il passaggio del flusso d'aria le *pliche vocali* si mettono in movimento vibrando, alternando cicli rapidissimi di chiusura e apertura causati dalla maggiore o minore pressione dell'aria messa in movimento dall'interno verso l'esterno.²⁴ Per questa ragione, quando le *pliche vocali* sono accostate, il flusso d'aria che proviene dai polmoni le mette in vibrazione, producendo un "rumore" che caratterizza i suoni *sonori*; quando queste sono aperte l'aria passa tra le due membrane senza provocare altro rumore se non un lievissimo fruscio con cui vengono prodotti i suoni detti *sordi*.

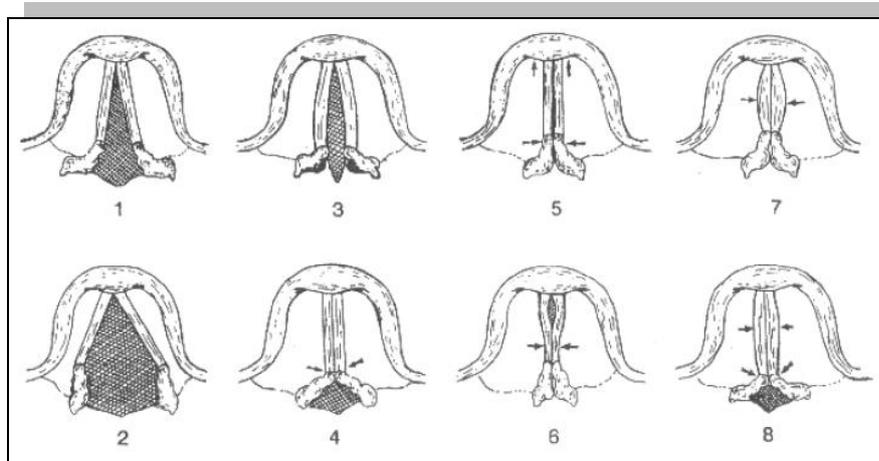


Figura 11: La glottide vista dall'alto nella posizione di 1) normale respirazione, 2) respirazione forzata, 3) produzione di foni sordi, 4) mormorio, 5) produzione di foni sonori, 6) laringalizzazione, 7) produzione dell'occlusiva glottidale, 8) bisbiglio²⁵

Il suono generato dalle vibrazioni delle *pliche vocali* viene successivamente modificato in vari modi con l'ausilio dei meccanismi *articolatorio* e *oro-nasale*, deputati alla vera e propria modificazione e differenziazione dei suoni generati dal meccanismo *glottidale* sotto la spinta dell'aria proveniente dal meccanismo *pneumatico*. Gli organi che intervengono in questa fase, responsabili delle differenze nei suoni prodotti dal parlante e percepiti dall'ascoltatore, possono essere raggruppati in organi *mobili* e organi *fissi*. Vengono definiti organi mobili le labbra, la mandibola, la lingua e le pliche vocali, chiamati anche *organi articolatori*, o semplicemente *articolatori* grazie alla loro capacità e libertà di movimento. Variando la posizione di questi elementi, il parlante modifica il flusso d'aria polmonare intervenendo sulla struttura delle cavità attraverso le quali passa (per esempio, cavità faringea, orale e nasale), creando delle vere e proprie casse di risonanza che differenziano, in base alla

²⁴ La frequenza con cui queste vibrano è misurabile e viene definita *frequenza fondamentale* (cfr. § 2.2.2).

²⁵ Figura adattata da Mioni (1986), pag. 71. La parte contrassegnata da colore scuro rappresenta il canale di passaggio del flusso d'aria proveniente dal meccanismo pneumatico.

loro conformazione, il suono che ne viene generato.²⁶ Sono invece organi fissi i denti, la radice dei denti, il palato duro e il palato molle, elementi che fungono in vari modi da supporto o “appoggio” agli articolatori mobili. In tal caso il suono subisce modificazioni in base all’entità della frizione impartita dagli articolatori mobili contro gli articolatori fissi restringendo o ostruendo il varco di uscita dell’aria. Ogni modificazione intervenuta e realizzata in questa fase al di sopra della glottide a carico dei due meccanismi articolatorio e oro-nasale in base ai contatti operati dagli articolatori mobili e fissi, produce cambiamenti e differenze nella qualità e nel tipo di suono che viene prodotto dal parlante e percepito dall’ascoltatore.

2.2.2 La frequenza fondamentale e le armoniche

L’apparato fonatorio è sostanzialmente uguale sia negli uomini che nelle donne e nei bambini, a parte alcune piccole differenze che rendono le voci chiaramente distinguibili l’una dall’altra, sia per i diversi sessi che per le varie fasce d’età. Queste diversità sono desumibili soprattutto dai valori di *frequenza fondamentale* (d’ora in poi F_0) che dipendono dalle caratteristiche anatomico-fisiologiche dell’individuo quali età, peso, sesso ed è direttamente legata alle caratteristiche delle pliche vocali quali elasticità, grossezza e lunghezza.

La *frequenza fondamentale* è la frequenza dell’onda generata dalla glottide, comunemente identificata con la vibrazione delle pliche vocali, e viene calcolata con un apposito algoritmo (*Cepstrum*) che consente di rilevare il suo valore anche in situazioni di segnale disturbato. Poiché le frequenze di un suono sono collegate tra di loro in modo armonico, il valore di F_0 corrisponde alla frequenza più bassa tra le componenti di un suono definite *armoniche*. Ogni armonica è multiplo intero di F_0 , pertanto la seconda armonica ha un valore di $2F_0$, la terza di $3F_0$ e così via. Il parlato è, in questi termini, il risultato dell’unione di più onde modificate dal nostro apparato fonatorio a formare onde complesse con F_0 e armoniche proprie e variabili. L’onda complessa che viene originata, infine, varia al variare dell’impostazione che l’apparato fonatorio assume nella produzione dei singoli foni.

La diversità maggiore tra uomo e donna sta nella differenza di altezza di F_0 (essa varia orientativamente da 80Hz a 200Hz per gli uomini e da 100Hz a 300Hz per le donne). L’uomo medio ha una laringe più larga, di conseguenza più adatta a produrre suoni con frequenza più

²⁶ Per questa ragione queste strutture vengono a pieno titolo considerate come dei filtri o dei risuonatori.

bassa rispetto alla donna media dotata di una laringe più stretta. Nei bambini questi valori sono ancora più alti rispetto agli adulti.²⁷

Strettamente connessi alla frequenza dell'onda generata dalla glottide, vi sono altri due parametri: *jitter* e *shimmer*. Per *jitter*, o perturbazione della frequenza, si intende la misurazione delle perturbazioni nella frequenza fondamentale tra cicli successivi.²⁸ Per *shimmer*, o perturbazione dell'intensità, si intende invece una misurazione della perturbazione dell'intensità, ovvero il calcolo della variazione dell'ampiezza della forma d'onda tra cicli successivi. Quelle appena presentate sono misurazioni ricavate con particolari algoritmi di calcolo e trovano per lo più applicazione in ambito diagnostico per quantificare e documentare eventuali patologie che influenzano e limitano le normali capacità di fonazione di un individuo. Entrambi i parametri variano infatti in funzione dei diversi tipi di fonazione (voce mormorata, rauca, sussurrata ecc.). Il calcolo di tali parametri viene solitamente effettuato su produzioni sostenute di foni vocalici (di norma la vocale /a/) che, per la loro conformazione spettrale relativamente stabile e ben definita (si veda quanto più sopra detto al riguardo), sono più idonee ad essere analizzate.²⁹ Per quest'ultima ragione, l'estrazione di tali parametri dal parlato continuo si rende particolarmente complessa poiché si assiste, in quest'ultimo, all'alternanza di foni vocalici e consonantici con tutta la variabilità e le problematiche che ne conseguono.

2.2.3 I suoni della parola

La conformazione fisica dell'apparato fonatorio risulta, come appena visto, essere determinante nella morfologia del segnale vocale. I suoni linguistici che grazie ad esso vengono realizzati, vengono comunemente divisi in due classi:

- **vocali**, ovvero suoni prodotti senza che via frapposizione di ostacoli al flusso d'aria nell'area che è compresa fra la glottide e il termine del percorso dell'aria, caratterizzate dalla presenza di vibrazione delle pliche vocali;
- **consonanti**, ovvero suoni prodotti con l'ausilio della frapposizione di un ostacolo, parziale o totale, al passaggio dell'aria in un qualche punto del percorso al di sopra della glottide.

²⁷ Per maggiori approfondimenti si rimanda a Whiteside (1996), Helfrich (1979), Braun (1995) e Chasaide Gobl (1997).

²⁸ Si tratta di una misurazione proposta da Lieberman (1961) come descrittore quantitativo della qualità della voce. La misura del *jitter* viene generalmente espressa in termini di stima percentuale (%) della differenza dei valori di F_0 tra cicli adiacenti.

²⁹ Per maggiori dettagli su questo aspetto si veda ad esempio il tentativo di Mendicino *et al.* (2009), di standardizzare alcuni parametri (tra cui anche *jitter* e *shimmer*) per applicazioni diagnostiche.

2.2.3.1 Le vocali

Le vocali risultano generate grazie all'emissione costante del flusso sonoro prodotto a livello glottidale. Poiché i suoni generati con concomitante vibrazione delle pliche vocali sono detti sonori, le vocali sono normalmente e per loro natura tutte sonore.

Le vocali sono in genere gli elementi più frequenti nel parlato e possiedono un'energia intrinseca maggiore rispetto a tutti gli altri foni. Per questa ragione le vocali risentono meno delle alterazioni e delle influenze indotte dal rumore. La loro differenziazione nella lingua parlata avviene sulla base di valori acustici detti *valori formantici* che sono strettamente connessi alla dimensione delle cavità sopraglottidali che fungono da casse di risonanza.

Le vocali possiedono, quindi, una tipica struttura spettrale caratterizzata dalla presenza di *formanti*. Come più volte detto, i suoni prodotti dal nostro apparato fonatorio sono caratterizzati dall'aver una determinata F_0 che domina su una serie di armoniche. Le formanti sono, perciò, dei rafforzamenti di energia di alcune armoniche in un determinato punto dello spettro. Questi picchi di energia corrispondono ai punti di maggiore annerimento sul tracciato sonografico (vedi Figura 6) o ai picchi dell'involuppo spettrale (vedi Figura 12).

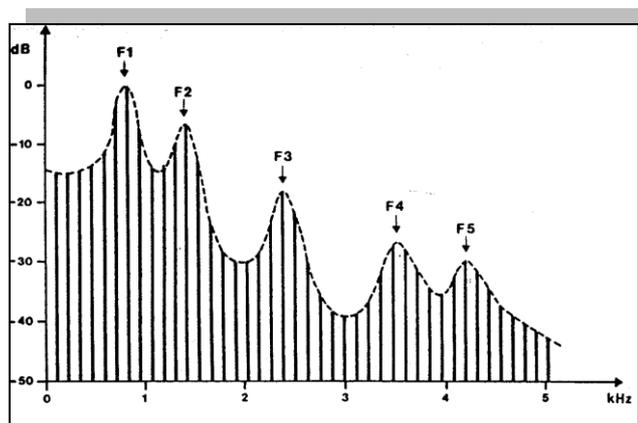


Figura 12: Spettro con involuppo spettrale delle prime cinque formanti di una vocale³⁰

Le formanti non coincidono necessariamente con un'armonica.³¹ Una formante può, infatti, contenere o dominare su più di un'armonica. Ne consegue che in assenza di armoniche da rinforzare non vi può essere formante che, per essere considerata tale, necessita di una determinata banda formantica o di un insieme di armoniche da rinforzare.³²

È necessario sottolineare, inoltre, che la diversificazione delle vocali da un punto di vista acustico-percettivo non dipende solo da una singola formante, ma dall'insieme di tutti i

³⁰ Figura tratta ed adattata da Ferrero *et al.* (1979).

³¹ Solo nel caso di un tono puro la frequenza formante coincide con l'armonica principale.

³² Mentre le armoniche sono per definizione multipli interi della frequenza fondamentale, le formanti lo sono raramente, ovvero solo in quei casi in il rafforzamento avviene in corrispondenza di una singola armonica.

valori formantici e dalla configurazione spettrale degli stessi. Ovviamente i valori formantici non sono sempre uguali, ma possono variare all'interno di un certo intervallo definito *area di esistenza vocalica*. Per questo motivo, nonostante la variabilità dei valori formantici, un ascoltatore sarà comunque in grado di identificarle.

La frequenza delle formanti³³ è differente dalla F_0 poiché dipende strettamente dalle cavità dell'apparato fonatorio, o meglio, dalle cavità che stanno al di sopra della glottide. Le cavità soprastanti hanno quindi funzione di filtri o di casse di risonanza che risuonano a determinate frequenze che, rafforzando una serie di armoniche danno origine a quelle che qui sono state definite formanti.

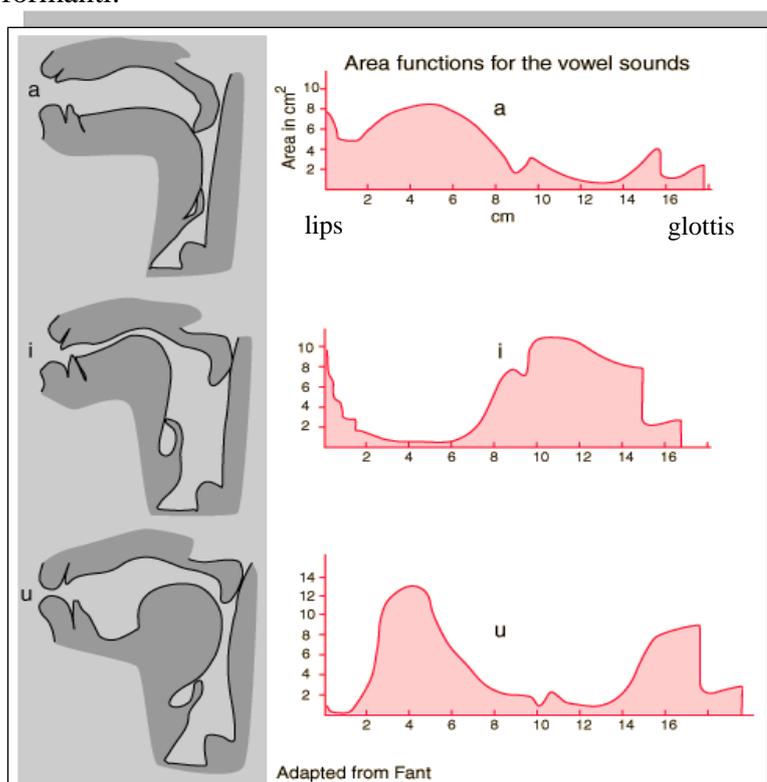


Figura 13: Schematizzazione del condotto vocale con relativa funzione d'area secondo Fant (1960)³⁴

Si tratta, in breve, di quella che è nota come la teoria acustica di produzione della parola attualmente e quasi universalmente riconosciuta, elaborata e sviluppata da Gunnar Fant (1960)³⁵, che approssima la configurazione del condotto vocale a un condotto acustico a

³³ Generalmente indicata con F_1 , F_2 , F_3 e così via.

³⁴ Fonte: <http://hyperphysics.phy-astr.gsu.edu/HBASE/music/fant.html> (data consultazione: 18 novembre 2007).

³⁵ La teoria proposta da Fant è anche nota come teoria sorgente-filtro in quanto "la parola è vista come il prodotto delle modifiche apportate al segnale - relativamente indifferenziato - emesso da una sorgente di suono, ad opera di un determinato filtro acustico complesso." Cfr. Ferrero, F., Genre, A., Boë, L. J., Contini, M. (1979), *Nozioni di fonetica acustica*, Torino, edizioni Omega, pag. 69.

simmetria radiale, asse rettilineo e sezione trasversa di area variabile: secondo tale teoria le frequenze formantiche rappresentano una funzione del condotto vocale.³⁶

Come già detto, la differenziazione delle vocali viene operata sulla base dei valori formantici, dove F_1 è inversamente proporzionale all'innalzamento/abbassamento della lingua (più la lingua è alta, più il valore di F_1 è basso; analogamente, più la lingua è bassa, più il valore di F_1 è alto); mentre F_2 , al contrario, è direttamente proporzionale all'avanzamento/arretramento della lingua (ciò significa che, più la lingua è anteriore, più F_2 è alta, mentre più la lingua è posteriore, più F_2 è bassa). Da quanto detto sinora, disponendo dei valori di F_1 e F_2 , sarà possibile individuare e definire le aree di esistenza vocalica per ciascuna vocale. Solitamente, per classificare le vocali, si utilizzano soltanto le prime tre formanti in quanto le altre formanti risultano essere troppo deboli, nella maggior parte dei casi impossibili da visualizzare e misurabili solo in condizioni di laboratorio.

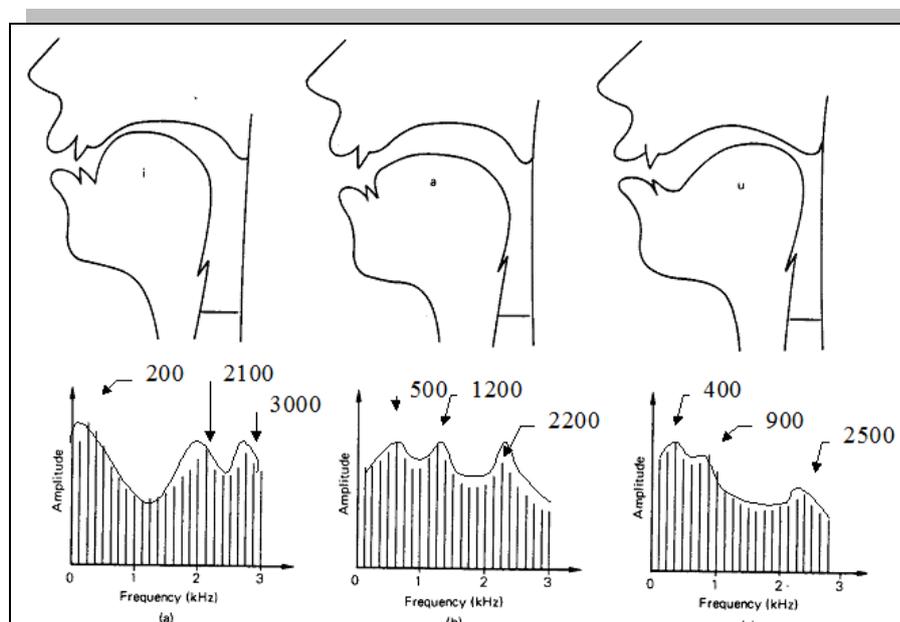


Figura 14: Classificazione delle vocali /i/, /a/ e /u/ sulla base delle impostazioni dell'apparato fonico-articolatorio con rispettivi spettri vocalici³⁷

Va inoltre ricordato che la produzione delle vocali, e quindi la loro composizione spettrale, viene fortemente influenzata dai foni che seguono o che precedono per il cosiddetto effetto di co-articolazione, dando origine a quelle che vengono definite *transizioni formantiche*.³⁸

³⁶ Cfr. *ibidem*, pp. 69-74. Le donne hanno generalmente un tratto vocale meno esteso, di conseguenza frequenze formantiche più alte. Vedi a questo proposito Diehl, Lindblom, Hoemeke, Fahey (1996).

³⁷ Figura tratta ed adattata da Romito (2000), pag. 127.

³⁸ Le transizioni formantiche consistono in deviazioni nella struttura formantica visibili su un tracciato sonografico e rappresentano la variazione della conformazione del condotto vocale dovuta al movimento degli articolatori mobili.

2.2.3.2 Le consonanti

Se una vocale viene definita come un suono risultante dal passaggio libero del flusso dell'aria attraverso le cavità orali, senza che vi sia attrito o impedimento, le consonanti vengono generate grazie ad interruzioni localizzate di tale flusso. Una consonante viene presentata come un suono che implica o un arresto (occlusione), che impedisce al flusso dell'aria di fuoriuscire (in tal caso si originano foni di tipo occlusivo), o una costrizione di questo flusso tale da farne percepire una certa frizione (foni di tipo fricativo). In base all'impedimento o al restringimento del passaggio del flusso d'aria, che può essere istantaneo o completo, si determina il modo di articolazione della consonante riconducibile alla disposizione degli articolatori mobili nella cavità orale. Poiché è abbastanza facile percepire l'avvicinarsi di un organo mobile ad uno fisso, oppure il loro contrario, da un punto di vista qualitativo i suoni consonantici sono più facili da descrivere di quelli vocalici. Contrariamente alle vocali che sono state definite come naturalmente sonore, le consonanti possono, invece, essere sia sonore che sorde, a seconda che vi sia o meno vibrazione delle pliche vocali.

Nella fonetica articolatoria, la composizione dei suoni consonantici, come anche la loro descrizione, viene indicata elencando il modo e il luogo di articolazione, rispettivamente occlusivo, nasale, *trill* o polivibrante, *tap* o monovibrante, fricativo, fricativo laterale, approssimante, approssimante laterale e bilabiale, labiodentale, dentale, alveolare, postalveolare, retroflesso, palatale, velare, uvulare, faringale, glottidale.³⁹

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)											
CONSONANTS (PULMONIC)											© 2005 IPA
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			ɾ					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Figura 15: Alfabeto Fonetico Internazionale (IPA) utilizzato per la rappresentazione grafica dei suoni consonantici⁴⁰

³⁹ Per una descrizione completa e più dettagliata di tali elementi consonantici, sia livello articolatorio che acustico, si veda Albano Leoni, F., Maturi, P. (2002), *Manuale di Fonetica*, Roma, Carocci Editore.

⁴⁰ Fonte: <http://www.langsci.ucl.ac.uk/ipa/index.html> (data consultazione: 13 settembre 2008). L'IPA è molto più ampio di quello qui riportato per le sole consonanti e prevede, oltre ad una serie di simboli per la rappresentazione di tutti gli altri suoni del linguaggio, comprese le vocali, anche una serie di diacritici per descrivere e rappresentare fenomeni di tipo soprasegmentale.

A livello grafico tutti i suoni del linguaggio vengono rappresentati con un apposito alfabeto, l'*International Phonetic Alphabet* (IPA). L'IPA (vedi Figura 15) è un alfabeto fonetico internazionale utilizzato per la trascrizione fonetica che, per convenzione, consente di attribuire in modo univoco uno specifico segno grafico a ciascun suono del linguaggio, risolvendo quelle che sono le differenze tra la comune grafia e i suoni concretamente prodotti.

Dalla Figura 15 si evince chiaramente quanto più sopra detto riguardo al modo e al luogo di articolazione leggibili, rispettivamente, nella colonna di sinistra e nella riga superiore.

A livello qualitativo le consonanti presentano sostanziali differenze dovute al loro modo di articolazione, differenze riscontrabili e misurabili a livello acustico.⁴¹

Da un punto di vista articolatorio le *occlusive* sono caratterizzate da tre fasi ben distinte che prendono il nome di *catastasi*, la fase in cui gli articolatori si preparano per il raggiungimento del target articolatorio; *tenuta*, la fase di occlusione o di restringimento del flusso d'aria; *metastasi*, la fase di rilascio e allontanamento repentino degli articolatori dal punto di occlusione. A livello acustico e spettroacustico la fase di *catastasi* non produce alcun effetto udibile o visibile. Al contrario, invece, ciò che comunemente viene percepito è il momento della *metastasi* che consiste in una vera e propria esplosione prodotta dal rilascio degli articolatori e che a livello spettroacustico risulta caratterizzato dalla presenza di *bursts* o *spikes* (striature verticali visibili sul sonogramma). In alcuni casi, a livello spettroacustico, si può anche rilevare, nel fono successivo a quello occlusivo, un ritardo dell'attacco dell'attività della laringe che viene indicato con il termine VOT (*Voice Onset Time*, o "ritardo dell'attacco della sonorità").⁴² Si tratta di un fenomeno (misurabile a livello temporale) contraddistinto dall'assenza di quella che sul tracciato sonografico viene definita *barra di sonorità* e che consente di distinguere a livello visivo di ispezione dello stesso, foni sonori da foni sordi. Per quello che concerne l'identificazione del luogo di articolazione delle occlusive sul tracciato sonografico ci si affida alle transizioni formanti che degli elementi vocalici attigui. A seconda del luogo di articolazione del fono occlusivo le transizioni formantiche (che rappresentano

⁴¹ Anche in questo caso, come già segnalato altrove in questo capitolo introduttivo, verranno di seguito forniti soltanto alcuni accenni agli elementi più salienti e generali utili a comprendere e a caratterizzare alcune differenze legate alla produzione dei diversi foni consonantici. Per questa ragione, data la superficialità e l'incompletezza delle informazioni fornite, per una trattazione più completa (che in questa sede si renderebbe assai prolissa) si rimanda ai più diffusi manuali didattici attualmente disponibili.

⁴² Quello del VOT è un elemento molto diffuso e presente soprattutto in lingue come l'inglese e il tedesco con caratteristiche di sistematicità. Nel tedesco può accadere, infatti, che in una sequenza che inizia per consonante sonora l'effetto del VOT faccia sì che l'attività laringea non inizi al costituirsi dell'ostacolo (*catastasi*), ma inizi con ritardo provocando una parziale o totale desonorizzazione della consonante stessa: per esempio nella parola *Bein* "gamba" [bæ:n].

l'adattamento della struttura delle cavità orali) tenderanno verso un punto o un valore di frequenza ideale definito *locus*.⁴³

Per quel che concerne le *fricative* bisogna distinguere tra sorde e sonore che a livello spettroacustico producono effetti diversi. Nelle prime, le sorde, si ha un annerimento verticale diffuso e continuo sul tracciato sonografico dato dalla frizione esercitata dagli articolatori mobili contro quelli fissi. È in base all'estensione della frequenza di tale rumore, che varia a seconda del luogo di articolazione, che si distinguono i diversi foni di tipo *fricativo*. Nel caso di foni *fricativi* sonori si ha, invece, oltre a quella che è stata definita *barra di sonorità*, una sostanziale riduzione degli effetti dell'attrito con una conseguente ed evidente riduzione di striature verticali sul sonogramma al posto delle quali si rileva, invece, una struttura formantica ben definita.

Le *affricate* possono, molto sommariamente, essere descritte come l'unione di un fono *occlusivo* e di un fono *fricativo* avente, per ciascuna delle sue componenti, le caratteristiche che sino ad ora sono state esaminate.

Le *nasali*, invece, presentano una struttura formantica molto simile a quella delle vocali e sono caratterizzate dall'aver una intensità più debole rispetto ad esse. La F1 dei foni nasali tende generalmente ad avere valori di frequenza alquanto fissi e stabili per ciascun parlatore (per effetto dell'abbassamento del *velo* si ha una parziale fuoriuscita dell'aria attraverso le cavità nasali che non sono modificabili), mentre la F2 presenta variazioni a seconda del luogo di articolazione del nesso nasale prodotto.

Le *vibranti* realizzate con repentine e ripetute interruzioni del flusso d'aria egressivo, si presentano costituite, a livello spettroacustico, dalla presenza di una barra di sonorità a cui si sovrappongono una serie di brevi fasi di silenzio (date dall'occlusione degli articolatori) seguite da brevissimi *spikes* (dati dal rilascio degli articolatori).

Per concludere, le consonanti *lateral*i vengono realizzate attraverso una parziale chiusura del flusso d'aria e presentano per questa ragione una struttura formantica molto simile a quella delle vocali, la cui intensità si trova a cavallo tra queste ultime e le consonanti nasali.

Le caratteristiche qui presentate sono chiaramente osservabili nel parlato *iperarticolato* (prodotto in laboratorio). Tuttavia, anche per le consonanti, come per le vocali, si verificano fenomeni di coarticolazione che ne complicano le indagini nel parlato continuo.

⁴³ Per maggiori dettagli e approfondimenti si veda Ferrero, F. E., Magno Caldognetto, E. (1986), Elementi di fonetica acustica, in *Trattato di Foniatria e Logopedia*, (L. Croatto, a cura di), Padova, La Garangola, pp. 172-178.

2.3 Cenni di Fonetica Uditiva e Percettiva

La Fonetica uditiva e percettiva è una ulteriore sottobranca della Fonetica che studia i processi di percezione dei suoni linguistici o foni. Come già evidenziato per la Fonetica articolatoria, anche la Fonetica uditiva e percettiva si basa essenzialmente su terminologie e conoscenze di altre discipline come l'anatomia, la fisiologia, la foniatria, la neurologia, l'otorinolaringoiatria, l'audiologia e non ultima la psicoacustica.

Si è più sopra fatto ricorso alla definizione di atto o processo comunicativo nei termini di una interazione audio-verbale tra due parlanti/ascoltatori A e B, ovvero un emittente e un ricevente. In questa sezione ci si occuperà, pertanto, del modo in cui i suoni che compongono il messaggio acustico-verbale del parlante-emittente colpiscono l'orecchio dell'ascoltatore-ricevente, degli effetti fisici che tali suoni esercitano su di esso e di come questi vengano percepiti e decodificati.

2.3.1 Fisiologia dell'apparato uditivo

L'organo dell'udito è il principale organo deputato alla ricezione e alla decodifica dei suoni linguistici. Contrariamente a quanto accade per gli organi coinvolti nella produzione dei suoni, la cui funzione è stata definita secondaria in virtù del fatto che nascono per assolvere ad altre funzioni primarie, l'organo dell'udito ha, invece, la funzione primaria dell'audizione, facoltà o abilità che risulta perfettamente sviluppata ed efficiente ancor prima della nascita. Il suo compito è quello di recepire vibrazioni acustiche e di convertirle in segnali atti ad essere trasmessi, lungo il nervo uditivo, alla parte del cervello che avrà il compito di analizzarli.

Le strutture dell'orecchio che costituiscono l'organo dell'udito sono rappresentate da una serie di canali, membrane, cavità e ossicini per mezzo dei quali i suoni o il messaggio verbale vengono convogliati e fatti arrivare alle strutture sensoriali vere e proprie situate nella parte più interna dell'orecchio, luogo in cui avviene la decodifica di tale segnale.

L'orecchio viene generalmente suddiviso in tre specifiche zone che sono l'*orecchio esterno*, l'*orecchio medio* e l'*orecchio interno* come riportato di seguito nella Figura 16.

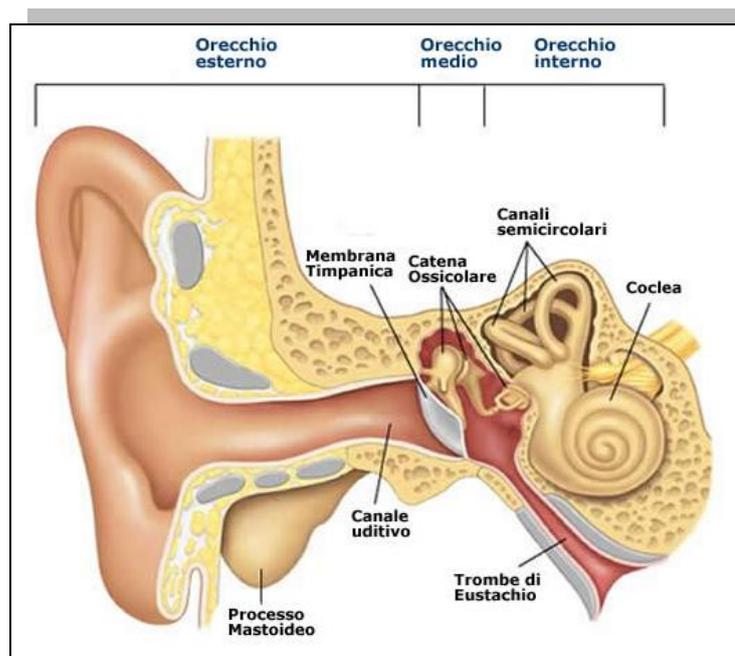


Figura 16: Sezione dell'orecchio umano⁴⁴

L'*orecchio esterno*, che ha una funzione secondaria nel processo uditivo, è costituito dalla parte più esterna e visibile rappresentata dal *padiglione* (a sua volta strutturato in *elice*, *anelice* e *lobo*) e dal *canale uditivo esterno*. Il *padiglione*, con la sua particolare conformazione, ha il compito di rinforzare il suono, ovvero di amplificarlo lievemente, e di farlo convergere verso l'interno dell'orecchio attraverso il *canale uditivo*. Il *canale uditivo* ha una lunghezza di circa 2,5 cm e termina, nella parte interna, con la *membrana del timpano* che viene colpita dalle onde sonore (che in questa fase sono ancora rappresentate da perturbazioni di particelle aeree) che verso di essa vengono convogliate. Il *timpano* è un organo sottile ed estremamente elastico, di forma circolare e sezione conica, che funge letteralmente da parete divisoria tra l'*orecchio esterno* e l'*orecchio medio*: ha il principale compito di trasformare le vibrazioni delle particelle aeree in vibrazioni di un corpo solido. La posizione del *timpano* all'estremità interna del *canale uditivo* protegge questa struttura da eventuali lesioni e contribuisce a mantenere costante ed invariata la temperatura e l'umidità circostante.

Al di là del *timpano* si trova l'*orecchio medio* (vedi Figura 17). L'*orecchio medio* è costituito da una cavità ossea, detta *cavo del timpano*, all'interno della quale si trova una catena di tre ossicini che, per la loro particolare forma, vengono chiamati *incudine*, *martello* e *staffa*. Il *cavo del timpano* è collegato con la faringe attraverso un canale che prende il nome

⁴⁴ Immagine tratta da: http://www.maicobioacustica.it/images/orecchio_normale.jpg (data consultazione: 16 novembre 2007).

di *tromba* o *tuba di Eustachio*.⁴⁵ Questi tre ossicini sono sospesi all'interno di questa cavità con un'assoluta libertà di vibrare ed hanno il compito di trasmettere gli impulsi del timpano ad un'altra membrana, la *finestra ovale*, su cui poggia la *staffa*, situata nella parte interna e superiore della cavità. In questa parte intermedia, situata tra l'*orecchio esterno* e quello *interno*, l'energia vibratoria viene ulteriormente accresciuta prima di giungere all'*orecchio interno*.⁴⁶ Sotto la *finestra ovale* si trova un altro foro, la *finestra rotonda*: queste due finestre mettono in comunicazione l'*orecchio medio* con l'*orecchio interno*, e comunicano tra di loro attraverso un canale lungo e stretto ripieno di un liquido, detto *perilinf*, che dalla *finestra ovale* corre verso l'interno attraverso la *scala vestibolare*. A metà strada questo canale torna indietro verso la *finestra rotonda* e prende il nome di *scala timpanica*.

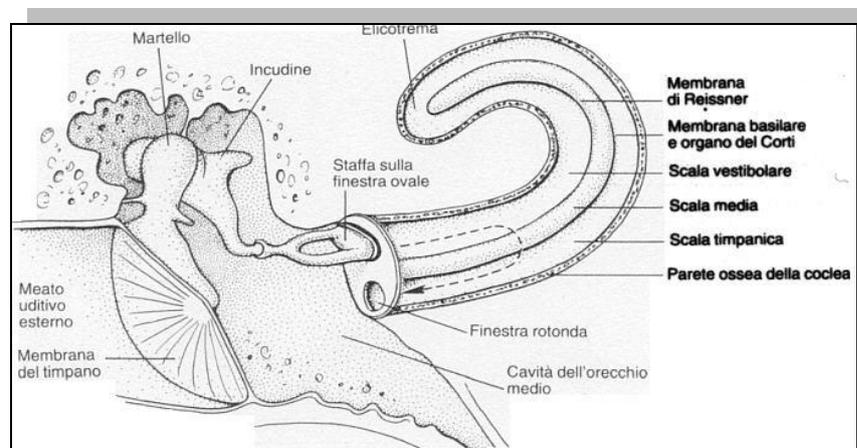


Figura 17: Sezione dell'orecchio medio⁴⁷

A questa parte dell'orecchio viene anche affidato il delicato compito di proteggere l'orecchio interno da suoni troppo intensi: tuttavia, la reazione degli organi coinvolti in questo

⁴⁵ Questo canale è normalmente chiuso e si apre quando si deglutisce o si sbadiglia favorendo scambi di aria allo scopo di uguagliare la pressione ai due lati della *membrana timpanica*. Questo meccanismo di stabilizzazione evita che il *timpano* venga deformato in modo brusco e doloroso. Un esempio concreto può essere quello di un raffreddore che ostruisce la tromba d'Eustachio costringendo il timpano a forti variazioni di pressione.

⁴⁶ Questo processo ha origine al livello del *timpano* con funzioni di trasduttore, che ha un'impedenza acustica molto bassa che favorisce una bassissima dispersione dell'energia del segnale che ad esso giunge. Per *impedenza* si intende un insieme di fattori che si oppongono al passaggio e alla trasmissione di un'onda sonora: si tratta più precisamente di una forza determinata dalle caratteristiche del mezzo (gassoso, liquido o solido) attraverso cui viene trasmessa l'onda ed è una misura della resistenza alla sua trasmissione. Rispetto ai gas, i liquidi mostrano una impedenza o resistenza maggiore alla pressione sonora. Quando le onde sonore viaggiano nell'aria (un mezzo gassoso) e raggiungono un liquido, molta dell'energia viene riflessa e solo parte di essa riesce ad attraversare il liquido. Poiché, come si vedrà, la *coclea* contiene un particolare liquido (*perilinf*), è necessario che la pressione sonora dell'onda che la raggiunge sia letteralmente aumentata per evitare fenomeni di dispersione di energia. In assenza dell'orecchio medio il suono trasmesso attraverso la perturbazione delle particelle dell'aria avrebbe notevoli difficoltà a raggiungere e ad essere utilizzato dalla *coclea*: si verrebbe infatti ad avere, a livello della *finestra ovale*, un contatto diretto fra un mezzo a bassa impedenza ed uno ad alta impedenza (*perilinf*) con una inevitabile ed elevata dispersione di energia.

⁴⁷ Immagine adattata da http://pcfarina.eng.unipr.it/dispense00/fossa124322/fossa124322_file/image038.jpg (data consultazione: 15 novembre 2007).

processo non è del tutto istantanea e suoni improvvisi possono provocare danni irreparabili all'udito.

Da qui, attraverso un canale lungo e stretto ripieno di *perilinf*a, si arriva all'*orecchio interno*, la struttura più interna dove si realizza quello che comunemente viene definito udito. All'interno di questa struttura hanno sede la *coclea*, l'organo dell'udito vero e proprio, il *nervo acustico* (o *uditivo*) e il *labirinto*, l'organo dell'equilibrio.

Tra la *scala vestibolare* e la *scala timpanica* corre, senza comunicare con questi ultimi, un altro canale, il *dotto cocleare* o *canale endolinfatico* immerso in un altro liquido, l'*endolinfa*⁴⁸. Nel suo insieme questa parte viene indicata con il termine *coclea* che, arrotolata su se stessa e per la sua struttura, ricorda quella del guscio di una chiocciola. Lungo tutta la parte interna del *dotto cocleare*, dalla parte della *scala timpanica*, si trova l'*organo del Corti*, l'organo sensoriale dell'udito, racchiuso tra la *membrana di Reissner* e la *membrana basilare* su cui è poggiato. L'*organo del Corti* si estende per tutta la lunghezza della *coclea* ed è avvolto a spirale come tutto il resto della struttura dell'*orecchio interno*.

L'*organo del Corti* è costituito da una serie di cellule, alcune con funzioni di sostegno (*cellule basali e di sostegno*), altre con funzioni sensoriali (*cellule ciliate*) di forma allungata e sottile fissate alla *membrana basilare*. Le *cellule ciliate* sono disposte su una fila di *cellule ciliate interne*, a cui spetta l'invio delle informazioni al *nervo acustico*, e tre file di *cellule ciliate esterne*, che hanno il compito di controllare l'ampiezza delle oscillazioni della *membrana basilare*. Al di sopra delle *cellule ciliate* si trova una lamina gelatinosa che prende il nome di *membrana tectoria* nella quale si inseriscono o si appoggiano le *cellule ciliate*.

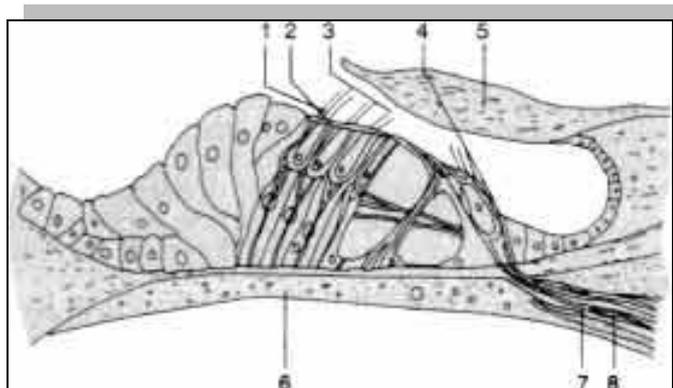


Figura 18: Sezione dell'organo del Corti con in evidenza 1) la membrana reticolare, 2) le cellule acustiche esterne, 3) le ciglia acustiche, 4) le cellule acustiche interne, 5) la membrana tectoria, 6) la membrana basilare, 7) la lamina ossea, 8) le fibre del nervo acustico⁴⁹

⁴⁸ La *perilinf*a differisce dall'*endolinfa* principalmente per la concentrazione degli elettroliti: la *perilinf*a è infatti più ricca in sodio e povera in potassio rispetto all'*endolinfa*.

⁴⁹ Immagine tratta ed adattata da <http://www.maurouberti.it/udito/udito5.jpg> (data consultazione: 15 novembre 2007)

In questa sede le vibrazioni trasmesse dalla *finestra ovale* al liquido *endolinfatico* raggiungono l'*elicotrema*, la parte terminale del dotto, per poi tornare indietro lungo la *scala timpanica* fino a giungere alla *finestra rotonda*. In questa fase le vibrazioni mettono in moto, piegandole, le *cellule ciliate*, situate lungo la *membrana basilare* che compone la *coclea*. Quando le *cellule ciliate*, che corrono parallelamente alle estremità delle fibre nervose del *nervo uditivo*, vengono messe in moto, le *ciglia* stesse, a contatto con la *membrana tectoria*, producono un'eccitazione elettrica che viene trasmessa alle fibre nervose sottostanti che, a loro volta, inviano il segnale ai neuroni del *nervo uditivo* per poi giungere alla corteccia uditiva, quell'area del cervello deputata alla ricezione delle sensazioni uditive.

Come è facile intuire, e come più sopra accennato, non tutti i suoni sono uguali. Poiché la *membrana basilare* aumenta di spessore e di elasticità man mano che ci si avvicina all'*elicotrema*, la stessa membrana genera, in risposta alle vibrazioni delle onde in arrivo, oscillazioni di ampiezza diversa in diverse zone a seconda della frequenza della onde: le frequenze più alte ecciteranno perciò le *ciglia* che si trovano nella parte iniziale della membrana, quindi più vicina alla *finestra ovale*; quelle di frequenza più bassa saranno captate nella parte più interna della membrana, ovvero verso l'apice o l'interno della *coclea*. Questa caratteristica della *coclea* viene anche detta *tonotopicità* (vedi Figura 19).

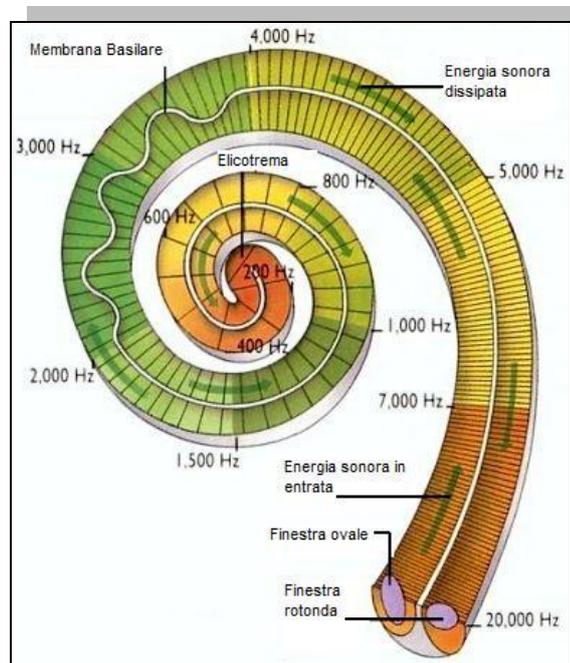


Figura 19: Tonotopicità della coclea con rappresentazione schematica della membrana basilare in funzione della frequenza⁵⁰

⁵⁰ Immagine tratta ed adattata da <http://universe-review.ca/I10-85-cochlea2.jpg> (data consultazione: 15 novembre 2007).

Quando a colpire l'orecchio sono suoni di tipo complesso, che nella vita di tutti i giorni sono una costante, è interessante notare come l'orecchio interno metta in atto una analisi spettrografica del suono in onde più semplici, un'analisi di *Fourier*, utilizzando una serie di filtri che altro non sono che le *cellule ciliate* situate nei vari punti lungo la *membrana basilare*.

2.3.2 La percezione dei suoni della parola

Al contrario di quanto si possa credere, la percezione dei suoni del parlato non avviene in base ad una pura e semplice decodifica dei suoni, che si susseguono l'uno all'altro nel *continuum fonico*, per ricavarne il messaggio in esso contenuto. Sebbene nella percezione del parlato emesso da un parlante l'ascoltatore operi una ricostruzione dei singoli elementi che compongono l'onda sonora che gli giungono all'orecchio, al fine di decodificare il messaggio trasmesso dal parlante risulta fondamentale e necessario il ricorso a tutta una serie di informazioni che derivano dalla conoscenza di determinate caratteristiche della lingua, oltre che della natura dei suoni stessi che la compongono.

Una delle più grosse difficoltà che contraddistinguono il parlato spontaneo è data dall'elevata velocità con cui le caratteristiche distintive dei singoli suoni vengono trasmesse.⁵¹ Come rilevato da Lieberman (1973), l'orecchio umano riceve qualcosa come 20-30 suoni al secondo, mentre è in grado di identificare correttamente solo i suoni che si trovano in una sequenza di 7-9 suoni al secondo (Lieberman, 1970). Questo significa che tutto ciò che giunge con una velocità pari, o superiore, a quella appena fornita viene percepito in un unico blocco rendendo complicando la possibilità di identificare i singoli suoni che compongono tale blocco. Stando così le cose risulta evidente che, se la percezione del parlato avviene comunque, questo non avviene certamente perché percepiamo ogni singolo suono.⁵²

In verità, l'uomo non è in grado di produrre o percepire un suono se non è seguito da un altro suono. Questo avviene soprattutto perché quello che è il target articolatorio, caratteristico dei singoli suoni della parola a cui più sopra si è accennato, in realtà, nel parlato quotidiano, non viene mai realizzato appieno. C'è piuttosto una tendenza da parte dei singoli segmenti a sovrapporsi e ad influenzarsi l'un l'altro all'interno di un'unica sequenza che

⁵¹ Ci si rende conto di questo aspetto soprattutto quando ci si ritrova di fronte ad una lingua che non si conosce o che si studia per la prima volta in cui di riflesso si chiede all'eventuale parlante straniero di "parlare più piano".

⁵² Un fatto analogo si verifica nella lettura dove non leggiamo le singole lettere, ma sequenze o blocchi di lettere, come ci si può rendere conto nell'esempio di seguito riportato: "Da uno stiduo dlel'Uvinrestia di Cmabirgde, non ha ipomrtanza in che odrnie le leterte sinoa in una proala, l'uicna csoa ipomrtatne è che la pimra e l'ulmtia lerteta sinoa al psoto guitso. Il rsteo può ervesse tolemtante sgliabato e si può lgegere sneza polbrmei. Qetsuo prhecé la mnete unama non lggee ongi lerteta ma la proala cmoe ucnio isenmie. Dveretinte, no?".

contraddistingue la sillaba come unità. Riprendendo l'esempio fornito da Matthei, Roeper (1991: 46):

“Così come non possiamo tenere separati i singoli gesti articolatori implicati nella produzione di una sillaba, non possiamo neppure isolare gli indici acustici prodotti da ciascun gesto articolatorio.”

Accade perciò, che gli indici e le informazioni della maggior parte dei nessi consonantici che precedono o che seguono una vocale vengono trasmessi alla stessa vocale che li separa (si veda quanto più sopra detto riguardo le transizioni formantiche e i *loci*)⁵³ grazie alla quale è possibile ricostruire il nesso consonantico che la precede o che la segue, mentre la vocale stessa viene riconosciuta come una specifica vocale se i suoi valori formantici ricadono all'interno di un determinato spazio acustico. A tutto ciò si aggiunge naturalmente l'esperienza linguistica dell'ascoltatore.⁵⁴ Ciò spiegherebbe, di fatto, l'elevata velocità che contraddistingue il parlato all'interno del quale l'informazione acustica trasmessa viene percepita in base ad unità della dimensione di una sillaba.⁵⁵

2.3.3 Teorie della percezione

Nel corso degli anni si sono succedute diverse proposte teoriche per spiegare e definire i meccanismi della percezione uditiva nell'essere umano. Sebbene si tratti di un ambito estremamente ampio e complesso che richiederebbe il dovuto approfondimento che in questa sede si rende impossibile, ci si limiterà solo ad un breve accenno alle teorie più diffuse.

Comunemente si suole suddividere tali teorie in teorie *attive* e teorie *passive*: le prime attribuiscono all'ascoltatore un ruolo attivo in cui lo stesso ascoltatore confronta ed elabora le caratteristiche oggettive del segnale con le forme immagazzinate nella memoria, affinché il riconoscimento del parlato possa realizzarsi solo per mezzo della mediazione del ricevente; di contro, le seconde attribuiscono all'ascoltatore un ruolo passivo, di semplice decodifica del segnale, segnale considerato di per sé sufficiente all'identificazione del contenuto del messaggio senza ulteriore intervento da parte dell'ascoltatore.

⁵³ Per i foni di tipo fricativo le caratteristiche acustiche intrinseche (di rumore della frizione e di intensità) sono di per sé già sufficienti al loro riconoscimento.

⁵⁴ Per quanto concerne esperimenti di percezione condotti in questo ambito, con test di *discriminazione* (tendenti a valutare la capacità di soggetti ascoltatori a percepire differenze tra vocali) e test di *identificazione* (allo scopo di identificare i confini percettivi tra due o più suoni vocalici lungo un continuum di stimoli) e per tutto ciò a cui qui non è stato fatto espressamente riferimento si veda Leoni, Maturi (2002), pp. 149-155.

⁵⁵ Cfr. Matthei, Roeper (1991), pp. 44-46.

Tra le teorie *attive* della percezione si ricorda qui la *teoria motoria* di Liberman (1967) e Liberman & Mattingly (1985), una delle più note, alla base della quale vi è l'assunto che il processo di percezione e il processo di produzione siano strettamente interrelati. Secondo questo approccio, chi riceve un segnale, opererebbe un confronto tra le caratteristiche acustiche decodificate dal sistema uditivo e le impostazioni articolatorie sottostanti alla produzione del segnale stesso. Solo dopo questa operazione di ricostruzione delle configurazioni articolatorie corrispondenti al segnale decodificato, l'ascoltatore sarebbe in grado di riconoscere la sequenza giuntagli all'orecchio. L'approccio descritto dalla teoria motoria è di tipo *top-down*, ovvero va da un livello alto relativo agli schemi neuromotori della produzione della parola ad un livello basso di riconoscimento di tratti e segmenti.

Tra le teorie *passive* si ricorda, invece, la teoria acustica di Fant (1960), che ha alla sua base il concetto di tratto distintivo.⁵⁶ Secondo questo approccio le strutture sensoriali dell'ascoltatore sarebbero in grado di riconoscere, nel segnale acustico, le caratteristiche fisiche che corrispondono ai diversi tratti distintivi, sufficienti di per sé a trasmettere tutta l'informazione relativa alla sequenza prodotta dal parlante. Partendo dalle caratteristiche distintive di ciascun suono percepito, il ricevente ricostruirebbe dapprima i singoli segmenti, poi le sillabe fino ad arrivare alle frasi. Da questo processo di ricostruzione che parte dalle singole unità per arrivare all'informazione nel suo complesso l'approccio viene anche definito di tipo *bottom-up*, ovvero dal basso verso l'alto. Secondo la teoria acustica di Fant, percezione e produzione sono da considerare due processi separati e paralleli, che operano in direzioni opposte.

Come parzialmente accennato nel precedente §, tuttavia, l'ascoltatore non fa ricorso soltanto alle informazioni scaturite dall'analisi del segnale, che potrebbero essere definite come informazioni interne. In quanto parlante di una comunità linguistica, un ascoltatore mette in atto un meccanismo di calcolo di previsione di ciò che sta ascoltando o che si aspetta di ascoltare sulla base delle conoscenze in suo possesso (conoscenza dell'interlocutore, del discorso, e di tutte quelle caratteristiche della lingua come lessico, sintassi, fonologia ecc.). Si tratta di elementi che possono essere definiti di informazione esterna e che nelle teorie della percezione sono state spesso trascurati, elementi che nella teoria H&H (*Hyper- and Hypo-speech*) di Lindblom (1987) risultano, invece, adeguatamente contemplati.⁵⁷

⁵⁶ I tratti distintivi, attribuiti a Roman Jakobson (1963), rappresentano una caratteristica binaria la cui presenza o assenza consente di distinguere tra loro i diversi fonemi (ad. esempio sonoro/sordo, nasale/orale, vocalico/non vocalico ecc.).

⁵⁷ Cfr. Leoni, Maturi (2002), pp. 158-159.

2.4 Nota conclusiva

In questo capitolo si è visto come quella che può essere definita una interazione audio-verbale tra due parlanti-ascoltatori A e B sia tutt'altro che semplice. Si tratta di un processo complesso che, come visto, chiama in causa diversi meccanismi sia da parte del parlante che dell'ascoltatore.

Senza andare oltre si chiuderà questa sezione con due rappresentazioni schematiche che, data la loro strutturazione, rendono pienamente conto di quelli che sono gli aspetti e i meccanismi della produzione e della percezione dei suoni del linguaggio qui presentati.

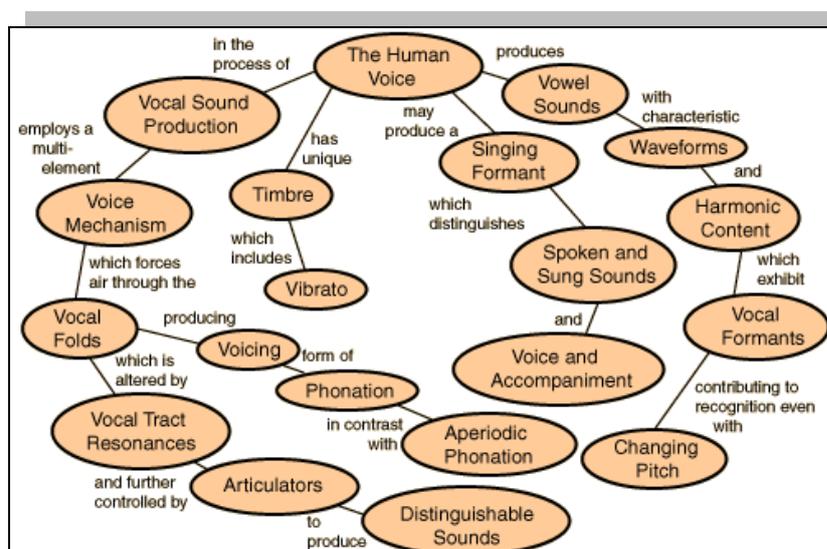


Figura 20: Meccanismi e caratteristiche della produzione dei suoni del linguaggio (vista parlante)

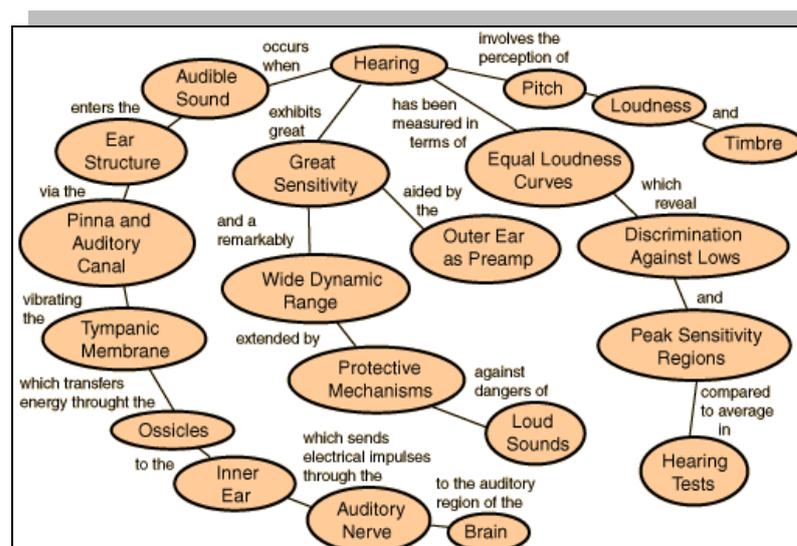


Figura 21: Meccanismi e caratteristiche della percezione dei suoni del linguaggio (vista ascoltatore)⁵⁸

⁵⁸ Fonti: rispettivamente <http://hyperphysics.phy-astr.gsu.edu/HBASE/music/voicecon.html#c1> e <http://hyperphysics.phy-astr.gsu.edu/HBASE/hframe.html> (data consultazione: 16 novembre 2007).

3 Le emozioni

Viviamo e proviamo emozioni ogni giorno della nostra vita e quotidianamente il nostro modo di interagire con gli altri e con l'ambiente che ci circonda viene regolato dalle stesse emozioni che viviamo: le emozioni influenzano e condizionano tutte le sfere della nostra vita, che vanno dal ragionamento ai processi decisionali, alla memoria, dalla percezione all'apprendimento, alla creatività. Ross (1981) dimostrò come la nostra abilità o capacità di esprimere emozioni sia strettamente legata alla nostra abilità di comunicare con gli altri.⁵⁹ Nello studio da questi condotto emerse come pazienti incapaci di identificare o produrre correttamente i tratti emotivi che accompagnano il parlato, incontrassero anche grosse difficoltà nei rapporti interpersonali, soprattutto nella capacità di esprimere se stessi.

Nell'interazione con gli altri, grazie all'immediatezza che contraddistingue sia la produzione che la ricezione di un atto comunicativo, l'essere umano è in grado di comprendere e di interpretare quasi istantaneamente e intuitivamente un qualsiasi messaggio e una qualsiasi emozione. Le emozioni determinano la nostra sopravvivenza, ci preparano all'azione, determinano il nostro carattere, regolano le nostre interazioni sociali, facilitano la comunicazione tra individui, facilitano l'interazione tra adulti e bambini, ci permettono di rispondere adeguatamente all'ambiente che ci circonda e cosa di non minore importanza, danno valore all'esperienza per la quale vale la pena di vivere a pieno la nostra vita.⁶⁰

Grazie alle caratteristiche appena elencate ciascuno di noi è in grado di capire se la persona con cui stiamo dialogando è triste o felice; allo stesso tempo ciascuno di noi, nel proprio piccolo, sa che cos'è la 'gioia', la 'tristezza', il 'disgusto', ma quanti di noi si sono

⁵⁹ È proprio in questo contesto che si inserisce il concetto di "intelligenza emotiva", il cui termine è stato coniato e introdotto quasi 20 anni fa da Salovey, P., Mayer, J. (1990), *Emotional Intelligence*, in: *Imagination, Cognition and Personality*, vol 9(3): 185-211. L'intelligenza emotiva è una capacità di discernimento che ci consente di riconoscere le nostre emozioni e quelle degli altri con l'obiettivo di organizzare e, di conseguenza, adattare la nostra vita e le nostre relazioni. Daniel Goleman, giornalista del *New York Times*, utilizzò, nel 1995, l'espressione "intelligenza emotiva" andando oltre la definizione proposta da Salovey & Mayer (1990) affermando che l'intelligenza emotiva è anche in grado di condizionare la nostra salute e i nostri successi (cfr. Oatley, K. (2007), *Breve storia delle emozioni*, ed. italiana a cura di L. Anolli, Universale Paperbacks il Mulino, Bologna, pag. 172).

⁶⁰ Vedi ad esempio De Sousa (1987: 332) che considera le emozioni alla pari del linguaggio e che alla pari del linguaggio caratterizza i contorni della nostra vita: "*Our emotional repertoires in some ways resemble our languages. Like language, emotion frames our possibilities of experience.*"

mai chiesti che cosa sia una “emozione”? Malgrado questa nostra esperienza, come testimonia una oramai nota affermazione di Beverly Fehr e James Russell (1984: 464), non siamo in grado di darne una vera e propria definizione:

“Everyone knows what an emotion is, until asked to give a definition.”

3.1 La terminologia delle emozioni

Nella nostra quotidianità usiamo termini come ‘emozione’, ‘stato d’animo’, ‘sentimento’ con estrema semplicità, anche se, come già accennato, ci risulta difficile darne una definizione nel momento in cui qualcuno ce lo chiede.

Prima di proseguire nella presente trattazione sarà quindi necessario fare il punto su quella che è la terminologia comunemente utilizzata in questo ambito di ricerca che, come si vedrà nelle pagine a seguire, risulta particolarmente variegato e tutt’altro che semplice.

Nell’affrontare la questione relativa alla terminologia è tuttavia indispensabile tenere sempre bene a mente quella che è la terminologia inglese utilizzata, terminologia che viene adottata a livello più generale e che, data la sua maggiore diffusione, lascia meno spazio a possibili fraintendimenti legati alla traduzione degli stessi termini.

Il termine ‘emozione’ (*emotion*), è forse il termine che in letteratura ha creato, e continua a creare, molti problemi, ma su questo aspetto si ritornerà in maniera più approfondita nel § 3.2. Per il momento sarà sufficiente richiamare la definizione riportata nel dizionario della lingua italiana Devoto-Oli (2008), dove, alla voce ‘emozione’ si legge:

*<e-mo-zió-ne> s.f.
Stato psichico affettivo e momentaneo che consiste nella reazione opposta dall’organismo a percezioni o rappresentazioni che ne turbano l’equilibrio; com., improvviso e forte turbamento provocato da commozione o da apprensione: gli ci volle non poco per rimettersi dall’e.; le e. della traversata; Avventura rischiosa ed eccitante: andare a caccia di emozioni. Dal fr. *émotion*, der. di *émouvoir* ‘mettere in moto, eccitare’ sec. XVII.*

Da questa definizione si evince chiaramente come le emozioni siano stati momentanei, caratterizzati da un *onset* improvviso di forte intensità e di brevissima durata che va incontro ad una fase di attenuazione.

Un altro termine molto ricorrente in questo ambito è quello di ‘umore’ o ‘stato d’animo’ (*mood*). In psicologia si è soliti differenziare tra *episodio emotivo* e *tratto emotivo*

come facce di una stessa medaglia. Sulla base di questa distinzione si tende a considerare l'emozione come *episodio emotivo* riferendosi sostanzialmente a:

“una risposta transitoria e momentanea a specifici eventi elicитanti, accompagnata da determinate modificazioni neuro psicofisiologiche, idonea a mantenere un adattamento adattivo all’ambiente”⁶¹,

mentre con *tratto emotivo* si tende ad identificare quello che comunemente viene definito stato d'animo, considerato:

“una disposizione o tendenza del soggetto a reagire prevalentemente secondo una definita modalità emotiva al flusso delle situazioni ambientali.”⁶²

Da questa distinzione emerge come ci si riferisca all'emozione come episodio (orientando l'attenzione sulla situazione che lo induce), mentre all'umore o stato d'animo come tratto (orientando l'attenzione sulla persona per giustificare le sue risposte emotive).⁶³

Un altro termine che viene spesso utilizzato come sinonimo di emozione è 'affetto' (*affect*) che deriva dal latino *afficere* e che vuol dire “portare qualcuno in una condizione organica o mentale, stimolare, impressionare”. In generale il termine 'affetto' sta ad indicare la qualità e i tratti centrali dell'esperienza emotiva ed attribuisce a quest'ultima un carattere non-cognitivo riferendosi essenzialmente alla valenza positiva (piacevole) o negativa (spiacevole) degli avvenimenti.⁶⁴ Il termine 'affetto' è pertanto un termine generico che può variare per qualità ed intensità e che comprende le emozioni. Tuttavia, sebbene tutte le emozioni siano 'affettive', non tutti gli stati affettivi sono emozioni.⁶⁵

Di natura più stabile è invece quello che comunemente chiamiamo 'sentimento' (*feeling*) che, come un'emozione, è rivolto verso un oggetto specifico. Il sentimento è frutto di esperienze precedenti o dell'apprendimento sociale e definisce quelli che sono i desideri, gli atteggiamenti e i comportamenti per trattare l'oggetto in una maniera corrispondente.⁶⁶ Allo stesso modo dello stato d'animo si tratta di uno stato affettivo di bassa intensità, durevole nel tempo la cui causa può anche non essere immediatamente percepibile.

⁶¹ Cfr. Anolli, L. (2002), *Le emozioni*, Milano, edizioni Unicopli, pag. 52.

⁶² *Idem.*

⁶³ Nella presente ricerca si farà quindi riferimento questo concetto di 'emozione'.

⁶⁴ Cfr. Anolli, L. (*op. cit.*), pag. 51.

⁶⁵ Cfr. Battacchi, M. W., Renna, M., Suslow, Th. (1995), *Emozioni e linguaggio*, Roma, La Nuova Italia Scientifica, pag. 9.

⁶⁶ Cfr. Anolli, L. (*op. cit.*), pag. 55.

Va inoltre fatto notare come spesso, tra i termini qui presentati, non vi sia un'esatta corrispondenza tra le lingue: ad esempio, mentre in italiano, francese e spagnolo il termine 'sentimento' sta ad indicare per lo più uno stato cognitivo, in inglese lo stesso termine rappresenta anche uno stato o una condizione fisica; al termine inglese 'emotion' in tedesco corrisponde il termine 'Gefühl' per indicare sentimenti a livello psicologico e fisiologico (vi è in questo caso analogia tra il tedesco 'fühlen' e l'inglese 'to feel'); o ancora, il termine inglese 'mood' in base al contesto in cui viene utilizzato può non corrispondere all'italiano 'stato d'animo', 'umore'.⁶⁷

3.2 Definire le emozioni

Riprendendo quanto affermato da Fehr & Russel (*op. cit.*) circa la nostra abilità a definire le emozioni, Cornelius (1996: 9) fa notare come la risposta a cosa siano le emozioni cambia a seconda del soggetto a cui lo si chiede e del momento in cui glielo si chiede. La questione della definizione delle emozioni è un problema in cui più o meno tutti gli studiosi si sono imbattuti e non è difficile scoprire come, a volte, gli argomenti addotti a supporto di questo aspetto manifestino un sottile velo di ironia, come ad esempio Young (1973), che suppone che tutti sappiano cosa siano le emozioni fuorché gli psicologi stessi.⁶⁸

Il primo a porsi la domanda "Che cos'è un'emozione?", fu William James (1842-1910), oggi considerato il padre della psicologia americana, quando nel 1884 scrisse un articolo per la rivista *Mind* dal titolo "What is an emotion?".

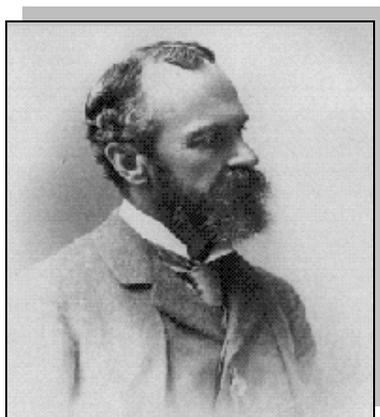


Figura 22: William James (ca. 1895), in: *The Letters of William James*, ed. by Henry James, Boston, 1920⁶⁹

⁶⁷ Per gli esempi appena riportati e per numerosi altri in lingue diverse si faccia riferimento ad Anolli, L. (*op. cit.*), pag. 123 e segg.

⁶⁸ Cfr. Battacchi, M. W., Renna, M., Suslow, Th. (*op. cit.*), pag. 17.

⁶⁹ *Op. cit.* in Goodman, R., "William James", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), <http://plato.stanford.edu/entries/james/> (data consultazione: 18 aprile 2008).

La risposta, che James diede all'epoca, fu che un'emozione è la percezione del cambiamento corporeo:

“My theory [...] is that the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion.”⁷⁰

Prima di lui, in epoche diverse e più lontane dalla sua, la discussione riguardo la questione delle emozioni non è stata certamente da meno. Vale qui la pena ricordare l'ampia ed esauriente esposizione di Solomon (1993) riguardo la “Filosofia delle emozioni”, in cui mette in evidenza come già gli antichi greci discutessero di emozione e temperamento, come anche del ruolo degli umori corporei nella regolazione delle emozioni, e come già Aristotele (384-322 a.c.), nella sua *Rhetorica*, definiva le emozioni,

“as that which leads one's condition to become so transformed that his judgement is affected, and which is accompanied by pleasure and pain. Examples of emotion include anger, fear, pity, and the like, as well as the opposite of these.”⁷¹

Ne scaturisce una concezione che vede nell'emozione uno stato perturbato/perturbatore in cui l'individuo è, in un certo qual modo, costretto a patire le emozioni.⁷² In epoca romana, nella filosofia degli Stoici, si riscontra, invece, una congiunzione/contrapposizione tra etica ed emozioni. Nella visione degli Stoici le emozioni rappresentavano errori concettuali e, come tali, venivano considerate fucina di miseria. Giungendo al Medioevo, lo studio delle emozioni resta ancora fortemente legato all'etica divenendo un punto centrale della psicologia Cristiana. Accanto ad elaborati studi sull'effetto dei vari “umori” sul temperamento dell'uomo⁷³, le emozioni vengono anche associate ai desideri che portarono il timore Cristiano per il peccato a minuziose analisi di particolari emozioni (passioni e desideri) etichettandole proprio come peccato. Che vi fosse in questo periodo un forte legame tra studio delle emozioni ed etica è dato anche dal fatto che virtù più alte come amore, speranza e fede

⁷⁰ Cfr. James, W. (1884), “What Is An Emotion?”, *Mind* 9: 190.

⁷¹ Cfr. Solomon (1993), pag. 4.

⁷² Gli antichi Greci chiamavano *pathèma* ciò che noi oggi chiamiamo emozioni e da cui deriva anche il termine *passioni*.

⁷³ Cfr. Irwin, J. R. (1947), Galen on the Temperaments, *Journal of General Psychology* 36: 45-64, *op. cit.* in: Scherer, K. R., Wranik, T., Sangsue, J., Tran, V. and Scherer, U. (2004), Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction patterns, *Social Science Information*, 43(4): 558. Per Galeno la tipologia di stimoli a cui l'individuo era esposto differenziava i quattro tipi di umore da lui postulati (melanconico, flemmatico, collerico e sanguigno), anche se l'origine dei temperamenti era, sempre secondo Galeno, da ricondurre ad uno sbilanciamento dei differenti fluidi corporei.

non venissero classificate come emozioni, ma venissero invece elevate ad uno status superiore, spesso identificato con la ragione.⁷⁴

Nonostante lo studio delle emozioni sia, come visto, storicamente radicato, il termine ‘emozione’ è un termine relativamente recente. La sua introduzione risale al XIX secolo, prevalentemente con accezioni di tipo scientifico e letterario al cui ceppo si fanno risalire anche parole come espressione, nervi, visceri e cervello. In passato si prediligevano, infatti, termini come passioni, sentimenti, sensazioni e non è un caso che tuttora i moderni dizionari affermino come l’emozione si trovi spesso in contrapposizione alla logica o alla razionalità o ancora alla cognizione.⁷⁵

Sul portale dell’Enciclopedia On-line della Treccani⁷⁶ alla voce ‘emozione’ viene data la seguente definizione:

“Processo interiore suscitato da un evento-stimolo rilevante per gli interessi dell’individuo. La presenza di un’e. si accompagna a esperienze soggettive (sentimenti), cambiamenti fisiologici (risposte periferiche regolate dal sistema nervoso autonomo, reazioni ormonali ed elettrocorticali), comportamenti ‘espressivi’ (postura e movimenti del corpo, emissioni vocali). Secondo le teorie funzionaliste sono il risultato di meccanismi adattativi in organismi che vivono in un ambiente complesso e incerto, e hanno la funzione di consentire all’individuo di reagire rapidamente, minimalizzando il processo di elaborazione delle informazioni, alle situazioni d’emergenza, sollecitando i comportamenti appropriati per affrontarle.”

Tuttavia, come sottolinea Solomon (1993: 4), dare oggi troppa enfasi al termine ‘emozione’ sarebbe un errore per il semplice fatto che:

“its range and meaning have altered significantly over the years, in part because of changes in theories about emotion.”

Ciononostante, quello delle emozioni è un ambito ampio, complesso e non ancora ben delineato che ha coinvolto, oggi più che in passato, innumerevoli scienziati che le hanno studiate dal proprio punto di vista privilegiato e con i mezzi che col trascorrere degli anni si sono a loro resi disponibili.⁷⁷ Data questa premessa è facile supporre come la questione del

⁷⁴ Cfr. Solomon (1993) a cui si rimanda per maggiori approfondimenti.

⁷⁵ Cfr. Oatley, K. (2007), *Breve storia delle emozioni*, edizione italiana a cura di Anolli L., Universale Paperbacks il Mulino, Bologna, pag. 117.

⁷⁶ Cfr. <http://www.treccani.it/Portale/homePage.html> (data consultazione: 14 marzo 2008).

⁷⁷ Si fa qui riferimento a mezzi sia tecnico-scientifici, come ad esempio particolari attrezzature, ma anche a nuove conoscenze e scoperte scientifiche.

“cosa” siano le emozioni rimanga un argomento ancora oggi ampiamente discusso e lontano dal trovare una risposta univoca sulla quale tutti concordino.⁷⁸

Il grande interesse che le emozioni hanno suscitato nei vari studiosi che se ne sono interessati, ha portato ciascuno di essi, come già aveva fatto James, a chiedersi che cosa effettivamente siano le emozioni dandone una propria interpretazione e definizione. A testimonianza di quanto appena detto basti dare uno sguardo al lavoro di Van Brakel (1993), nel quale vengono riportate ben 24 definizioni di “emozione” rintracciabili in varie pubblicazioni dal 1980 in poi e che, di seguito, vengono riportate in ordine cronologico rispetto alla fonte originale da cui sono state tratte.

Averill (1980, 1982)	<i>Emotions are transitory social roles that include an individual's appraisal of the situation, are interpreted as a passion rather than as an action and are responses that have been institutionalised by society as a means of resolving conflicts.</i>
Lazarus et al. (1980)	<i>Emotions are complex, organised states consisting of cognitive appraisals, action impulses, and patterned somatic reactions.</i>
Leventhal (1980)	<i>Emotion words are related to (but not referring to) certain bodily feelings; they are intentional (about something) and they involve a moral order.</i>
Lyons (1980)	<i>X is to be deemed an emotional state if and only if it is a physiologically abnormal state caused by the subject of that state's evaluation of his or her situation.</i>
Tomkins (1980)	<i>Affects (like "joy", "fear", "angry") are sets of muscular and glandular responses located in the face and also widely distributed throughout the body.</i>
Marks (1982)	<i>Emotions are belief/desire complexes characterised by strong desire.</i>
Hochschild (1983)	<i>Emotion is a biologically given sense (like hearing, touch, smell) which is related not only to an orientation toward action, but also toward cognition.</i>
Dienstbier (1984)	<i>Human emotion is a motivation-laden feeling resulting from shifts in arousal and the meaning attached to those arousal shifts.</i>
de Rivera (1984)	<i>Emotions are concerned with adjusting the relationship between person and other, ideally maximising the value of the relationship.</i>
Lang (1984)	<i>An emotion is defined as a broad response disposition that may include measurable language behaviour, organised overt acts, and a physiological support system.</i>
Levy (1984)	<i>Emotions are about one's mode of relationship as a total individual to the social and nonsocial environment.</i>
Lutz (1984)	<i>Emotions are culturally constructed concepts which point to clusters of situations typically calling for some sort of action.</i>
Frijda (1986)	<i>Emotion is a script-like process. It starts with eliciting conditions which have a range of attributes. It continues with evaluations in relation to the individual's concerns. It causes physiological disturbances, and prepares the</i>

⁷⁸ Seppur dichiaratamente pessimista, le emozioni continuano ancora oggi a stimolare il confronto, lo studio e la ricerca dando vita ad ulteriori approfondimenti e lasciando ampio spazio di azione a quanti si occupano di emozioni.

	<i>individual for action. Finally, it issues in expressions and actions.</i>
Harré (1986)	<i>Emotions are culture-specific patterns of learnt behaviour, mirroring a society's conception of what is appropriate in a particular, socially defined, situation.</i>
de Sousa (1987)	<i>Paradigm scenarios, in setting up our emotional repertoire, quite literally provide the meaning of our emotions; their axiological correctness is defined in terms of these scenarios.</i>
Koch (1987)	<i>Emotions share something with sensations, perceptions, beliefs, desires, and moral judgements. They have phenomenal, intentional, and evaluative aspects and have to do with interactive complexes of bodily feelings, attentional sets, evaluations or (moral) judgements, and desires/impulses/inclinations to behave in certain ways.</i>
Greenspan (1988)	<i>Emotions are affective mixtures of comfort/discomfort, inhibition/arousal, etc. towards certain evaluative propositions spelling out their intentional content.</i>
Solomon (1988)	<i>An emotion is an evaluative (or normative) judgement, a judgement about my sensation and about myself and/or about all other people.</i>
Johnson & Oatley (1989)	<i>There is a small number of basic signals that can set up characteristic emotional modes within the organism. Terms that refer to basic emotions cannot be analysed into anything more basic, such as a prototype or a set of semantic features. Other terms refer to states that combine a basic emotion with a propositional content.</i>
Middleton (1989)	<i>Emotions are sociocultural creations that are necessarily communicated routinely through a variety of channels serving to clarify intentions, attitudes, identity and meaning.</i>
Ratner (1989)	<i>Emotions depend on a social consciousness concerning when, where, and what to feel as well as when, where, and how to act.</i>
McCullagh (1990)	<i>Emotions involve some sort of reaction to an evaluation, usually at least the onset of feelings about it, sometimes desires and physiological changes as well. These reactions are not rational, but automatic responses, and therefore emotions cannot be wholly rational.</i>
Oakley (1992)	<i>Emotions are complex phenomena involving dynamically related elements of cognition, desire, and affectivity. All three of these elements have moral significance. Emotions are essential and enduring features of our moral character.</i>
Shweder (1993)	<i>Emotions are complex narrative structures which give shape and meaning to somatic and affective experiences whose unity is to be found in the type of self-involving stories they make it possible for us to tell about our feelings.</i>

Tabella 2: Definizioni proposte per le emozioni in varie pubblicazioni dal 1980 in poi⁷⁹

Analizzando attentamente le definizioni sopra riportate, è evidente come ciascuna delle definizioni si rifaccia ad uno specifico costrutto teorico, facendo emergere un fatto innegabile giustamente evidenziato da Solomon (1993: 3):

⁷⁹ Tabella tratta ed adattata da Van Brakel (1993), tabella 9. Ove non diversamente segnalato vale quanto riportato dallo stesso autore, ovvero “*Descriptions given are literal quotations or paraphrases from sources given*”.

“Just when an adequate definition is in place, some new theory rears its unwelcome head and challenges our understandings.”

Come riporta lo stesso Van Brakel, qualche decennio prima di lui, Plutchik (1980) aveva già individuato e contato 28 definizioni di emozione. Ma Van Brakel e Plutchik non sono certamente stati gli unici a cimentarsi in una simile iniziativa. Kleinginna & Kleinginna (1981: 345),

“In an attempt to resolve the resulting terminological confusion”,

sono andati oltre quella che potrebbe risultare una semplice raccolta di definizioni, classificando ben 92 definizioni e 9 affermazioni scettiche in 11 categorie sulla base dei fenomeni emozionali o presupposti teorici in esse enfatizzati.⁸⁰ Concludendo il loro lavoro, dopo aver esaminato e commentato più di 100 definizioni di emozione, Kleinginna & Kleinginna (1981: 355) suggeriscono che:

“a formal definition of emotion should be broad enough to include all traditionally significant aspects of emotion, while attempting to differentiate it from other psychological processes.”

Inoltre, poiché un accordo sulle tante caratteristiche che distinguono le emozioni è assai difficile, Kleinginna & Kleinginna (*ibidem*) propongono una definizione che enfatizza i molti e possibili aspetti delle emozioni, ovvero:

“Emotion is a complex set of interactions among subjective and objective factors, mediated by neural-hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-directed, and adaptive.”

In termini analoghi a quelli proposti da Kleinginna & Kleinginna si è espressa anche Balconi (2004: 31) definendo le emozioni come:

⁸⁰ All'interno questa della raccolta sono ovviamente presenti anche le 28 definizioni contate da Plutchik (1980), come anche le 11 contate da Fantino (1973) da cui Kleinginna & Kleinginna (1981) sono partiti per la loro indagine. Dato il numero delle definizioni e la lunghezza della lista, si rimanda alla fonte per un elenco completo delle definizioni oggetto di studio degli autori.

“[...] esperienza che produce sentimenti o affetti (‘mi sento felice’), presenta delle modificazioni fisiologiche in risposta a determinati stimoli (un aumento del battito cardiaco, ad esempio), è dotata di correlati cognitivi (come la valutazione degli stimoli emotigeni) ed ha, infine, risvolti sul piano del comportamento.”

Leggermente più dettagliata, anche se non molto diversa da quelle riportate, è la definizione di emozione data da Scherer (1986: 146):

“a series of interrelated adaptive changes in several organismic subsystems following antecedent events evaluated to be of major relevance to an organism’s goals.”

Col passare degli anni, e a distanza di decenni, si denota, infine, una tendenza a dare maggiore peso e importanza a quelli che Balconi chiama “risvolti sul piano del comportamento”, ovvero a considerare le emozioni come fenomeni di tipo sociale, aspetto su cui si ritornerà nel § 3.3.3 con le teorie costruttivistiche. Ne sono una testimonianza le definizioni di emozione che ne danno Oatley (2007), per il quale le emozioni sono:

“i mezzi fondamentali con cui si strutturano i rapporti umani. (pag. 110) [...] avvengono a livello individuale e come tali possono essere osservate e misurate. (pag. 188)”

o Mesquita (2007: 413), per la quale:

“Emotions are interpersonal events that navigate the social context and are informed by it. All emotional events (perhaps everything more complex than a startle) are intensely meaningful, and meaning is provided by the socio-cultural context.”

Non bisogna tuttavia dimenticare che le emozioni sono anche fenomeni di tipo soggettivo in quanto esse rappresentano:

“[...] la componente soggettiva, la sensazione affettiva, che accompagna la condotta di un individuo.”⁸¹

Dalle definizioni sin qui proposte, a cui se ne potrebbero aggiungere infinite altre, emerge come siano numerosi e differenti i gruppi di studiosi che si interessano di emozioni (che spaziano dalla biologia alla genetica e alle neuroscienze, dalla psicologia alla sociologia e all’etologia, dall’economia alle scienze comportamentali, dalla fisica alle scienze della voce

⁸¹ Cfr. Canestrari, R., Godino, A. (2000), *Psicologia generale, le strade maestre della conoscenza*, Milano, Paravia-Bruno Mondadori editore, pag. 210.

e della comunicazione). Nonostante esistano e siano state proposte così tante definizioni di emozione, Scherer (1994) sottolinea come uno dei motivi legati allo scarso avanzamento delle ricerche sulle emozioni sia da ricercarsi nella frammentazione del fenomeno stesso, all'interno del quale, oltre alla complessità metodologica insita in certe indagini e l'alta specializzazione che queste richiedono,

“the consequence seems to be a serious lack of agreement on the definition of emotion and a dearth of collaborative efforts to study different facets of emotional experience and behavior”.⁸²

Ciò che si può senza dubbio affermare è che le emozioni costituiscano un particolare punto di contatto fra aspetti biologici, soggettivi, sociali e culturali che permeano fino in fondo l'esperienza umana, attribuendo significato agli eventi quotidianamente vissuti.

3.3 Teorie sulle emozioni

Dall'elevato numero di definizioni proposte nel precedente paragrafo, che meriterebbero senz'altro di essere analizzate una per una, è facile capire come ciascuna di esse faccia riferimento, per quanto possibile, ad una specifica teoria piuttosto che ad altre.

Un esempio di ciò che sono le teorie delle emozioni potrebbe essere dato da un semplicissimo esempio identificabile nel racconto buddista dell'elefante e del gruppo di non vedenti, considerato un esempio classico nella psicologia delle emozioni. Si tratta della storia di un gruppo di uomini non vedenti in viaggio che, improvvisamente, si imbatte in un grosso elefante. Tutti nel gruppo hanno sentito parlare di questa enorme creatura, ma nessuno ne ha mai incontrata una. Di conseguenza, ciascuno di loro cerca di definire che cosa effettivamente questa creatura sia o a che cosa assomigli ma, a causa della loro disabilità legata alla mancanza della vista, ciascuno di loro ha un'idea differente. Il primo di loro ne studia la zanna e decide che un elefante è molto simile ad una lancia. Un altro ne studia accuratamente il corpo e alla fine decide che l'elefante è simile ad un grande muro. L'ultimo ne studia attentamente la coda e in conclusione decide che un elefante assomiglia ad una fune.

L'esempio riportato dovrebbe in questo caso mettere in evidenza come un determinato oggetto, l'elefante nel racconto, le emozioni nel caso qui esaminato, possa essere studiato da diverse prospettive, da diversi punti di vista, producendo, al tempo stesso, visioni diverse

⁸² Cfr. Scherer, K. R. (1994), Affect Bursts, in: *Emotions: Essays on Emotion Theory*, (S.H.M. van Goozen, N.E. VandePoll, & J. A. Sergeant, Eds.), Hillsdale, NJ: Erlbaum, pp. 161.

dello stesso oggetto di studio. È infatti ciò che emerge dalle diverse teorie delle emozioni che sino ad oggi sono state formulate.

Alcune teorie si concentrano di più su alcuni aspetti delle emozioni, altre su altre componenti ancora. Ciascuna teoria guarda alle emozioni con gli strumenti e le conoscenze che la disciplina di appartenenza gli mette a disposizione. Una esemplificazione di quanto accade con le emozioni può essere data attraverso un diagramma di Venn (vedi Figura 23), nel quale viene messo in evidenza come l'oggetto di studio, le emozioni appunto, possano essere osservate e studiate da diverse prospettive.

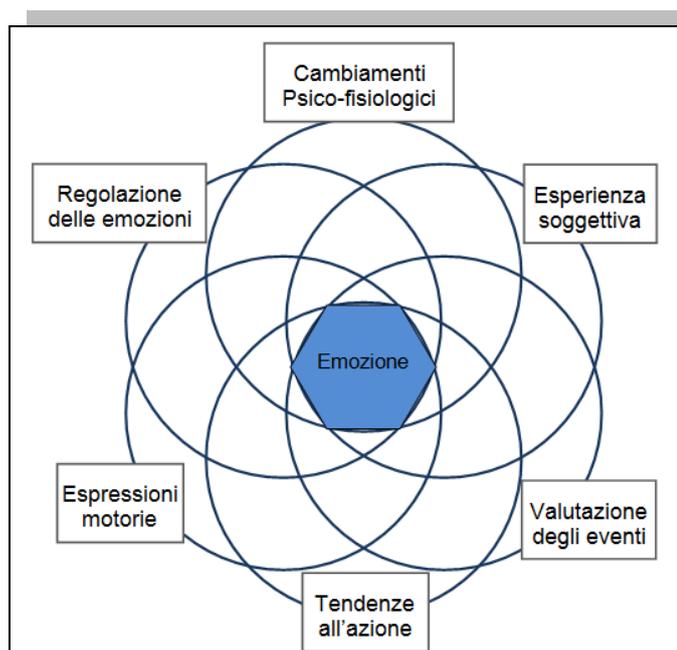


Figura 23: Esempificazione dei tipi di approccio allo studio delle emozioni a partire da punti di vista particolareggiati dettati dalle diverse discipline scientifiche

Citando ancora una volta Van Brakel (1993),

“Various approaches to emotions have been proposed in philosophy, biology, psychology, and social science at large. Just as with emotions, classifications of theories of emotions vary.”

A tal proposito, come riportato in Tabella 3, Cornelius (1996: 12) identifica quattro tradizioni teoriche riconducibili rispettivamente alla tradizione Darwiniana, Jamesiana, Cognitiva, Sociale-Costruttivistica.

Tradition	Key Idea	Classic Work	Contemporary Research
Darwinian	Emotions have adaptive functions, are universal	Darwin (1872/1965)	Ekman <i>et al.</i> (1987)
Jamesian	Emotions = bodily response	James (1884)	Levenson <i>et al.</i> (1990)
Cognitive	Emotions are based on appraisals	Arnold (1960)	Smith and Lazarus (1993)
Social Constructivist	Emotions are social constructions, serve social purposes	Averill (1960)	Smith and Kleinman (1989)

Tabella 3: Le quattro tradizioni teoriche delle emozioni in psicologia identificate da Cornelius⁸³

Ciononostante, riuscire a fornire una classificazione precisa e chiara di quelle che sono le correnti di pensiero o le teorie legate alle emozioni è alquanto difficile perché gli stessi ambiti di studio a cui queste fanno riferimento non possiedono, a loro volta, confini ben delimitati.

Per quelle che sono le tematiche qui affrontate, ci si limiterà perciò ad un accenno alle singole teorie secondo una tripartizione delle teorie delle emozioni secondo la proposta di Scherer, Johnstone, Klasmeyer (2003) in teorie dell'*arousal*, teorie delle emozioni discrete e teorie dell'*appraisal*, con l'aggiunta di una quarta categoria, quella delle teorie costruttivistiche.

3.3.1 Teorie dell'*arousal*

Sotto questa etichetta vengono identificate tutte quelle teorie che, quale in misura maggiore e quale in misura minore, sanciscono l'importanza dell'attivazione biologica (*arousal*)⁸⁴ dell'organismo nella definizione e nell'espressione delle emozioni.

Una delle prime teorie delle emozioni postulate in questi termini è quella conosciuta sotto il nome di James-Lange sul finire dell'800. Si è già fatto cenno a quella che fu la risposta di James alla domanda "Che cosa sono le emozioni?", la cui risposta fu appunto "che i cambiamenti corporei seguono direttamente la percezione dell'evento emotigeno, e che la percezione di tali cambiamenti nel momento in cui si manifestano produce l'emozione".⁸⁵ Le emozioni scaturiscono perciò dalla percezione dei cambiamenti neurofisiologici che si

⁸³ Tabella tratta ed adattata da Cornelius (1996), pag. 12, Table 1.1.

⁸⁴ Da qui anche l'etichetta di *arousal theories*.

⁸⁵ Tale teoria, conosciuta anche come teoria periferica o fisiologica (proprio perché si basa sulle attività del sistema nervoso periferico che governa nervi sensoriali e motori), viene postulata da James 12 anni dopo la pubblicazione del trattato di Charles Darwin "*The Expressions of the Emotions in Man and Animals*", nel 1872. È proprio sotto l'impulso delle scoperte e dell'opera di Darwin che ha inizio un secolo all'insegna della ricerca scientifica in numerosissimi ambiti scientifici. La teoria prende anche il nome di Carl Lange, un fisico danese che, in modo del tutto indipendente, propose una teoria molto simile intorno al 1885.

manifestano. Stando a questa teoria, per James è sbagliata l'idea, ancora oggi assai diffusa, che nel vedere un orso ci spaventiamo e scappiamo. Il punto, secondo James, sta nel fatto che vediamo l'orso e scappiamo e di conseguenza abbiamo paura dell'orso.⁸⁶ Si evince, da questa concezione, come la paura sia la percezione dell'attivazione del nostro corpo che, stimolato dalla situazione di emergenza, mette in atto la fuga: in altri termini, i nostri sistemi sensoriali inviano informazioni al cervello sulla situazione attuale e, in risposta, il cervello invia segnali al corpo, specificamente al Sistema Nervoso Autonomo, provocando cambiamenti del tono muscolare, del battito cardiaco, dilatazione della pupilla, ovvero si generano una serie di meccanismi di difesa che sono innati, per dare origine, in termini di percezione del cambiamento, all'emozione.⁸⁷

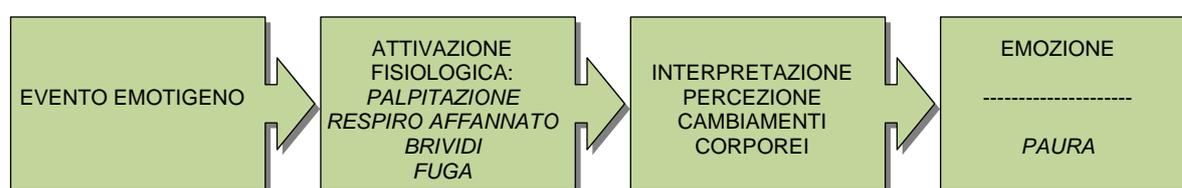


Figura 24: Schematizzazione della teoria James-Lange

Secondo la concezione alla base di questa teoria, rimuovendo le modificazioni fisiologiche che sono all'origine dell'emozione, si dovrebbe poter dire che la stessa emozione svanisce con essi. A supporto di tale teoria, infatti, esiste uno studio condotto da Hohmann (1966) che intervistò pazienti, con danni in diversi punti della colonna vertebrale, sulla differenza di intensità tra emozioni vissute prima e dopo l'incidente. I risultati riportati da Hohmann indicano che, maggiore è il grado di infermità prodotto dalla lesione, maggiore è la diminuzione delle sensazioni, nel caso specifico, di paura e odio. Questi dati sembrerebbero avvalorare, in conclusione, l'idea che per provare emozioni sia necessario avere percezione delle reazioni fisiologiche che si verificano nel proprio corpo.⁸⁸

Nel 1927, Walter B. Cannon, fisiologo americano, propone una visione antitetica a quella appena descritta, dimostrando come la teoria proposta da James-Lange sul funzionamento del corpo umano fosse da ritenere non valida. A supporto di questa sua

⁸⁶ Analogamente noi non piangiamo perché siamo tristi, piuttosto ci sentiamo tristi perché piangiamo; non tremiamo perché abbiamo paura, ma abbiamo paura perché tremiamo; siamo felici perché sorridiamo, e così via.

⁸⁷ Cfr. LeDoux, J. E. (1996), *The Emotional Brain*, New York: Touchstone, pag. 43.

⁸⁸ Questa ipotesi sembra essere confermata da ulteriori studi, come, ad esempio, quelli condotti da Damasio (1994: 94) che descrive pazienti affetti da *anosognosia* che sono incapaci di esperire emozioni (si tratta di soggetti che non sono in grado di percepire il proprio corpo a causa di una paralisi che coinvolge la parte sinistra del corpo dovuta ad una lesione delle aree somato-sensoriali dell'emisfero destro). Se nel secolo scorso la teoria di James-Lange è stata in un certo qual modo discreditata, vi sono altri studiosi come John Cacioppo, Antonio Damasio, Joseph E. LeDoux e Robert Zajonc che hanno rivalutato alcuni assunti teorici postulati da James, grazie e soprattutto alla possibilità che questi ultimi hanno avuto nell'avvalersi di prove scientifiche di natura neurologica.

concezione Cannon porta cinque prove per confutare il fatto che i cambiamenti corporei sono alla base dell'esperienza emotiva:

1. La totale separazione delle viscere dal sistema nervoso centrale non altera il comportamento emotivo: cani e gatti, le cui viscere erano state separate dal sistema nervoso centrale, erano ancora in grado di manifestare emozioni, il che comporterebbe la possibilità di esperire un'emozione anche se non è possibile avvertire i cambiamenti fisiologici;
2. Le stesse modificazioni viscerali si manifestano in stati emotivi differenti e in situazioni non emotive: Cannon sosteneva, infatti, come molti cambiamenti corporei fossero simili per molti tipi di emozioni, mentre la teoria James-Lange presupponeva la risposta di differenti *pattern* corporei per emozioni diverse;
3. Le viscere sono strutture relativamente insensibili a causa della loro scarsa innervazione;
4. I cambiamenti a livello viscerale sono troppo lenti per essere considerati fonte di sensazione emotiva;
5. L'induzione artificiale di modificazioni viscerali tipiche di forti emozioni non le riproduce.⁸⁹

Secondo la teoria di Cannon, la regione talamica, in presenza di uno stimolo indotto da un evento emotigeno, invia impulsi simultaneamente alla corteccia cerebrale, producendo la sensazione consapevole dell'emozione, e al sistema nervoso simpatico responsabile dell'attivazione fisiologica: i centri di attivazione, controllo e regolazione dei processi emotivi sono perciò localizzati centralmente nella regione talamica, in quanto i segnali nervosi sarebbero in grado sia di indurre le manifestazioni espressivo-motorie delle emozioni, sia di determinare le loro componenti attraverso le connessioni con la corteccia celebrale.⁹⁰

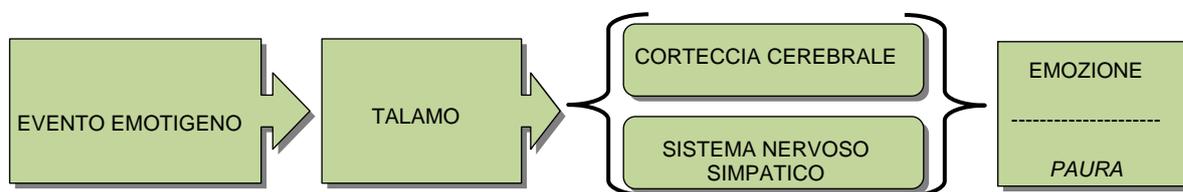


Figura 25: Schematizzazione della teoria di Cannon

⁸⁹ Cfr. Cornelius, R. R. (1996), *The science of emotion. Research and tradition in the psychology of emotion*, Upper Saddle River, NJ: Prentice-Hall, pag. 58 e segg., in cui viene fornita un'analisi molto dettagliata circa le motivazioni addotte da Cannon contro le idee di James sulle emozioni come percezione delle modificazioni corporee. A supporto di questo ultimo punto Cannon utilizzò i dati di uno studio sugli effetti dell'iniezione di adrenalina nella percezione delle emozioni pubblicate, nel 1924, da Gregorio Marañón, un fisico spagnolo (cfr. *ibidem*, pag.73).

⁹⁰ Data l'attribuzione dell'origine delle emozioni al sistema nervoso centrale, la teoria di Cannon viene anche identificata come teoria centrale (in contrapposizione a quella periferica di James-Lange).

La teoria Schachter-Singer (1964) può invece essere considerata come la fusione della teoria di James e della teoria di Cannon con l'aggiunta di una componente di tipo cognitivo⁹¹: per percepire un'emozione sarebbe necessaria una valutazione cognitiva dell'attivazione corporea intervenuta, lasciando in tal senso ampio spazio all'interpretazione soggettiva delle emozioni. I due psicologi propongono, infatti, una visione nella quale gli individui spesso esperiscono modificazioni o attivazioni fisiologiche che non sempre hanno una causa nota o una giustificazione. Tali fenomeni fisiologici portano e motivano l'individuo a ricercare una spiegazione alla causa e alla natura dell'attivazione (o *arousal*) che, di conseguenza, dà origine ad uno stato emotivo evidente.

Dalle proposte di Cannon e di Schachter-Singer emergerebbe, quindi, come l'origine delle emozioni sia di tipo cognitivo piuttosto che di tipo viscerale. Tuttavia, sebbene venga solo parzialmente riconosciuta l'importanza della valutazione cognitiva come preconditione per il manifestarsi delle emozioni, per avere una presa di posizione più forte ed articolata su tale aspetto bisogna rifarsi alle idee formulate da Arnold (1960) e suoi successori per quelle che vengono oggi riconosciute come *appraisal theories*.⁹²

3.3.2 Teorie delle emozioni discrete

Quelle che qui vengono identificate con l'etichetta di teorie delle emozioni discrete possono essere ricollegate alla visione delle emozioni in termini di evoluzione della specie, che attribuiscono alle emozioni caratteristiche di innatezza o di universalità. Si tratta, per lo più, di teorie scaturite e sostenute da studi sulle espressioni facciali che, seguendo le orme e le teorie di Darwin, hanno portato ad assumere l'esistenza di un set di emozioni universali (e quindi innate), che si possono chiaramente distinguere l'una dall'altra e che rappresentano per questa ragione la base di tutte le altre emozioni che l'uomo è in grado di esperire.

La teoria delle emozioni discrete trova fondamento nelle ricerche condotte da colui che viene oggi considerato il pioniere degli studi sull'universalità delle emozioni, Charles

⁹¹ Per l'accostamento della componente cognitiva alla componente fisiologica, la teoria è anche conosciuta come teoria dei due fattori (*Two-Factor Theory*) o ancora cognitivo-attivazionale. La teoria era stata formulata in termini simili 40 anni prima dal fisico spagnolo Marañón, dal quale lo stesso Cannon aveva ripreso alcuni dati sperimentali a supporto della sua teoria, e al quale si sono rifatti anche i due psicologi Schachter e Singer riprendendone alcuni studi. Tuttavia, nonostante le prove addotte a sostegno di questa teoria risultino alquanto deboli (e privi di valenza statistica), l'idea che la cognizione fosse in grado di differenziare l'esperienza emotiva di un individuo ha comunque avuto grande influenza nella formulazione di numerose altre teorie, soprattutto quelle basate sull'*appraisal* (cfr. Niedenthal, P. M., Krauth-Gruber, S., Ric, F. (2006), *The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches*, Principles of Social Psychology series, New York: Psychology Press, pag. 14). Nonostante venga in questo caso chiamata in causa la componente cognitiva, l'argomento centrale della teoria proposta resta comunque l'*arousal* corporeo.

⁹² Sulle posizioni caratterizzanti di queste teorie si ritornerà più approfonditamente nel § 3.3.4.

Darwin (1809-1882), autore de *L'Espressione delle Emozioni nell'Uomo e negli Animali* (1872).

Darwin fu il primo studioso a sottolineare il ruolo determinante delle emozioni nel processo di evoluzione: la convinzione interculturale dell'esistenza di emozioni umane innate risale, infatti, alle indagini da lui stesso condotte nel 1872 sull'osservazione di espressioni vocali e facciali simili di specifiche emozioni di persone di diversa cultura.⁹³ Dalle osservazioni condotte sugli animali trovò, allo stesso, tempo una continuità filogenetica, suggerendo per le emozioni l'esistenza di evidenti fondamenti biologici ed evolutivi. Contrariamente a Bell e Duchenne, che ricorrevano alla figura di un creatore per spiegare l'universalità delle espressioni facciali, Darwin ne fece, invece, un valido strumento per sostenere l'ipotesi evuzionista.⁹⁴ Ne *L'Espressione delle Emozioni nell'Uomo e negli Animali* prevale, da questo punto di vista, la tesi secondo cui le emozioni e le rispettive espressioni facciali sono di tipo adattivo, originatesi dall'evoluzione della specie. Questa caratteristica le renderebbe, quindi, biologicamente innate ed universali e, in quanto tali, "patrimonio genetico" di ogni essere umano escludendo, di fatto, l'idea che le emozioni rappresentino un qualcosa di culturalmente appreso. È in base a tale principio che nasce l'assunto che tutti gli esseri umani esprimano, attraverso il volto, le stesse emozioni nello stesso identico modo prevaricando quelle che possono essere definite distinzioni di razza o cultura.

Tra la pubblicazione di Darwin e i successivi studi sulle espressioni facciali passeranno parecchi anni che, a parte qualche sparuto studio, hanno lasciato sull'argomento un *vacuum* scientifico non indifferente dovuto, sostanzialmente, all'impostazione Skinneriana e al dettame comportamentista secondo il quale le emozioni erano semplicemente inosservabili: venivano addirittura considerate prive di interesse ai fini della ricerca psicologica in quanto le espressioni facciali erano considerate gesti arbitrari appresi in infanzia attraverso la "punizione" e la "ricompensa".⁹⁵

⁹³ Anche se, ad oggi, pare che i primi a parlare di universalità delle espressioni facciali siano stati Bell (nel 1806), seguito da Duchenne (nel 1862).

⁹⁴ Si tratta di quella che Oatley K. (2007), *Breve storia delle emozioni*, edizione italiana a cura di Anolli L., Universale Paperbacks il Mulino, Bologna, pag. 13, definisce concezione evuzionista delle emozioni, facendo risalire alcuni aspetti delle nostre emozioni a milioni di anni fa, al tempo in cui i nostri antenati (non ancora umani) iniziarono a distinguersi rispetto ai membri di altre specie pre-umane che non erano in grado di sopravvivere perché meno dotati di quelle risorse emotive e pratiche che, invece, erano in possesso dei nostri antenati: risorse ed elementi vincenti del passato che ci sono state tramandate attraverso i geni.

⁹⁵ I pochi studi condotti in questo periodo riguardano, prevalentemente, aspetti psico-fisiologici a cui si è già fatto cenno nel precedente §.

L'interesse per le espressioni facciali si infervora verso la metà degli anni '60 grazie a Sylvan Tomkins (1962; 1963) e ai suoi allievi Paul Ekman e Carroll Izard che portarono avanti una serie di studi oggi conosciuti con il nome di *universality studies*.

Gli studi furono condotti in diverse culture e la valutazione di diverse espressioni sulla base di foto statiche portò alla convinzione e alla dimostrazione dell'esistenza panculturale di una serie di emozioni, quelle che oggi conosciamo col termine di *basic emotions* o emozioni universali o ancora primarie e che comprendono rabbia, disgusto, paura, gioia, tristezza e sorpresa.⁹⁶

A tali studi seguirono una serie di ricerche condotte da parte di Ekman & Friesen (1971) su persone appartenenti a culture pre-letterate a supporto della tesi sull'universalità delle espressioni facciali.

Ulteriori ricerche da parte di Charlesworth & Kreutzer (1973) dimostrarono che le medesime espressioni erano riscontrabili in primati non umani ed in individui nati ciechi.

Come già detto, si tratta di teorie che si rivolgono quasi esclusivamente alle emozioni delle espressioni facciali e non è un caso che, nell'ambito delle ricerche sul parlato emotivo, una delle maggiori accuse rivolte ai sostenitori di questa teoria viene da Scherer (2003: 244) che afferma:

“Unfortunately, to date there has been no attempt by discrete emotion theorists to develop concrete predictions for vocal patterning.”

3.3.3 Teorie costruttivistiche

I sostenitori di questo filone teorico, tra i quali Averill (1980), Harré (1986), Mandler (1990), ritengono che le emozioni siano dei costrutti sociali che vengono appresi e condivisi a livello culturale. In quest'ottica le emozioni non sono fenomeni o processi naturali preposti alla sopravvivenza biologica, ma soddisfano esigenze di tipo sociale regolando a vari livelli le relazioni che intercorrono tra gli individui. Sia l'espressione delle emozioni, sia le emozioni stesse, vengono costruite a livello culturale, incluse le esperienze soggettive. Le emozioni

⁹⁶ Sull'esistenza di un set di emozioni limitate, o su quelle che possono essere considerate le *basic emotions* si è instaurato un ampio dibattito: per maggiori dettagli o approfondimenti si rimanda al volume a cura di R. J. Davidson, P. Ekman (1994), *The nature of emotion: Fundamental questions*, New York: Oxford University Press, in cui alcuni tra i più autorevoli studiosi in questo ambito si sono posti la domanda “Are there basic emotions?”. Ad ogni modo, Ekman (1971) suggerisce l'esistenza di 6 emozioni *basic* (*happiness, anger, fear, sadness, disgust, surprise*); Izard (1971) suggerisce che ve ne siano 8 che vengono interpretate in modo simile in contesto cross-culturale (oltre a quelle identificate da Ekman aggiunge *interest e shame*). Come suggeriscono Magno Caldognetto, E. *et al.* (2008), La faccia e la voce delle emozioni, in: *La mente del cuore*, a cura di I. Poggi, Armando Editore, Roma, pag. 179, è sulle 6 emozioni *basic* di Ekman che è stato raccolto il maggior numero di dati e a cui nella maggior parte delle ricerche condotte si fa riferimento.

sono una specie di comportamento sociale appreso con l'educazione attraverso l'acquisizione di norme che regolano il comportamento degli individui in determinate circostanze. Si tratta di una visione che sta in netta contrapposizione alle teorie che considerano le emozioni come programmi determinati a livello genetico e quindi innati. Attraverso la cultura e la socializzazione vengono apprese quelle che sono le regole delle emozioni, in base alle quali l'individuo è in grado di gestire e organizzare le proprie esperienze emotive. Nella teoria costruttivista le emozioni hanno una funzione prescrittiva: un insieme di risposte e comportamenti socialmente prescritte che l'individuo utilizza in determinati contesti e che hanno un determinato significato all'interno del sistema socioculturale in cui si manifestano.⁹⁷ Si tratta di regole diverse dalle regole di esibizione (*display rules*) di Ekman (1982) che sono, invece, regole culturalmente determinate e socialmente condivise che prescrivono come, quando e in quali situazioni un'emozione primaria (un'emozione molto intensa a livello biologico) può essere o meno espressa.

Infine, i sostenitori della teoria costruttivista, pur riconoscendo l'esistenza di fondamenti biologici delle emozioni, ritengono l'importanza di questi ultimi come secondaria rispetto ai meccanismi e alle relazioni socialmente costruite attraverso cui le emozioni vengono modulate.⁹⁸

3.3.4 Teorie dell'*appraisal*

Anche sotto questa etichetta, come nei casi precedentemente illustrati, si registrano diverse posizioni teoriche. In generale, sotto la categoria di "teorie dell'*appraisal*" vengono identificate tutte quelle teorie recenti, spesso anche modelli, che riconducono le emozioni e la loro differenziazione al risultato di tutta una serie di processi cognitivi⁹⁹ che Clore & Ortony (2000: 32) hanno definito come:

"psychological representations of emotional significance".

Sebbene le posizioni di Cannon e Schachter-Singer contemplassero l'intervento di una componente cognitiva per la valutazione dell'attivazione corporea all'origine delle emozioni, è con le idee di Arnold (1960), a cui si era già parzialmente accennato nel paragrafo 3.3.1, che si assiste a quello che potrebbe essere definito il momento fondante delle moderne teorie

⁹⁷ Cfr. Anolli, L. (2002), pp. 46-48.

⁹⁸ Per una più esauriente trattazione si rimanda a Cornelius (1996).

⁹⁹ Cfr. Scherer, Johnstone & Klasmeyer (2003), pag. 435.

cognitive delle emozioni. Secondo la Arnold (1960: 175) l'*appraisal*, alla quale viene attribuita la coniazione del termine, rappresenta:

“the direct, immediate sense judgment [...]”

una valutazione diretta, immediata ed allo stesso tempo avente natura non-riflessiva, non-intellettuale ed automatica¹⁰⁰, senza la quale non vi possono essere emozioni. Per evocare una emozione:

“the object must be appraised as affecting me in some way, affecting me personally as an individual with my particular experience and my particular aims.”¹⁰¹

È evidente come la Arnold insista sulla valutazione cognitiva per l'organizzazione dei comportamenti emotivi in cui l'esperienza soggettiva influenza il modo di valutare le situazioni e gli eventi. Dall'enfasi che la Arnold pone sull'*appraisal*, definendo l'emozione come:

“The felt tendency toward anything intuitively appraised as good (beneficial), or away from anything intuitively appraised as bad (harmful). This attraction or aversion accompanied by a pattern of physiological changes organized toward approach or withdrawal. The patterns differ for different emotions.”¹⁰²

si denota come vengano contemporaneamente colmate quelle che erano le lacune della teoria Jamesiana. In tal senso è il processo di valutazione a dare origine ai cambiamenti fisiologici, alle sensazioni e alle espressioni tipiche di ciascuna emozione. Volendo rappresentare l'idea della Arnold in modo schematico la sequenza all'origine delle emozioni è la seguente:



Figura 26: Schematizzazione teoria dell'appraisal di Arnold (1960)

Uno dei primi a seguire e ad approfondire quelle che furono le idee della Arnold fu Richard Lazarus. Per Lazarus (1966) l'emozione è un fenomeno post-cognitivo all'origine della quale vi è una fusione tra funzioni cognitive altamente sviluppate e meccanismi di valutazione con cui viene elaborato e definito il significato di uno stimolo emotivo. Il concetto di *coping* rappresenta uno dei concetti cardine della sua teoria e con esso viene definito

¹⁰⁰ Cfr. Arnold (1960), pag. 175.

¹⁰¹ *Ibidem*, pag. 171.

¹⁰² *Ibidem*, pag. 182.

l'insieme dei processi e delle modalità con cui vengono elaborati e valutati gli stimoli provenienti dall'ambiente circostante. Nel quadro teorico proposto da Lazarus vi è una differenziazione delle componenti cognitive con l'individuazione di due tipi di *coping*. Il primo comprende azioni dirette e urgenti per far fronte a situazioni di imminente minaccia e pericolo in cui viene valutata la situazione (*primary appraisal*) come anche la capacità di far fronte alla situazione (*secondary appraisal*). Entrambi i processi di *appraisal* viaggiano parallelamente per dare origine ad una nuova valutazione (*reappraisal*) che è, invece, un processo meramente cognitivo che non implica necessariamente il ricorso ad azioni, ma mette in atto un'analisi del successo o del fallimento che le azioni dirette hanno determinato. In questo continuo processo di valutazione e rivalutazione all'attività di *coping* viene affidata la funzione di regolare gli scambi tra l'ambiente e l'individuo e quindi la qualità e l'intensità delle emozioni esperite.¹⁰³

Il modello proposto da Ortony, Clore, Collins (1988) e conosciuto anche come *OCC model*, rappresenta un altro punto di vista tra quelli che sono stati qui raggruppati sotto il nome di *appraisal theories*. Nel modello *OCC* le emozioni sono considerate come risposte a tre tipi di stimoli: eventi (interpretazioni che le persone danno delle cose che accadono); agenti (cose o in genere persone che possono causare o contribuire al verificarsi di un evento); oggetti. Per gli eventi l'elemento centrale alla valutazione cognitiva è dato dalla loro piacevolezza con riferimento a quelli che sono i propri scopi o obiettivi (*pleased/displeased*); per gli agenti è invece data dall'apprezzabilità delle azioni con riferimento agli standard (*approve/disapprove*); per gli oggetti, infine, è data dalla loro appetibilità (*like/dislike*) con riferimento alle proprie attitudini. Le emozioni sono assimilabili ad una reazione a catena che ha sempre inizio con una reazione positiva o negativa in risposta agli eventi, ad azioni degli agenti o a qualche aspetto degli oggetti: possiamo essere contenti o dispiaciuti per le conseguenze di un evento; possiamo approvare o disapprovare l'azione di una persona; possiamo apprezzare o non apprezzare un oggetto. L'emozione si basa, quindi, sull'esperienza e sulle sensazioni soggettive. Il meccanismo di valutazione cognitiva non è necessariamente conscio (una persona può esperire una sensazione di paura senza però essere in grado di stabilire su quale conoscenza questa sensazione sia basata). La funzione dell'emozione è

¹⁰³ Cfr. D'Urso, Trentin (2004), pag. 97. Va inoltre fatto notare come negli anni ottanta il dibattito sulla relazione tra emozione e cognizione sia stato animato da un acceso scontro tra Lazarus (1982, 1984), che sosteneva il primato della cognizione nella reazione emotiva e l'interdipendenza dei due processi, e Zajonc (1980, 1984) che sosteneva, invece, che emozione e cognizione fossero due processi separati e solo in parte dipendenti dimostrando come si potessero esperire emozioni anche in assenza di processi cognitivi. Per maggiori dettagli su questo dibattito si veda Battacchi, Renna, Suslow (1995), pag. 33 e segg.

quella di focalizzare l'attenzione sull'evento emotivo e preparare l'individuo all'azione qualora il caso lo dovesse richiedere.¹⁰⁴

Il modello *OCC* offre una rappresentazione e strutturazione gerarchica degli stati emotivi basandosi su quelli che vengono considerati gli elementi che inducono una determinata emozione da cui scaturisce la valutazione cognitiva. Tale aspetto lo rende particolarmente interessante in sistemi di Intelligenza Artificiale in quanto ne rende possibile l'implementazione sia a livello teorico che computazionale. Diversi sistemi di agenti conversazionali hanno adottato il modello ricorrendo ad una sorta di modulo definito "affective reasoning": in base al dominio della conoscenza, una situazione viene valutata ricorrendo al modello *OCC* determinandone la reazione emotiva più appropriata, influenzando le conseguenti azioni ed il comportamento espressivo dell'agente.¹⁰⁵

Maggiormente in linea con le tematiche qui affrontate troviamo, infine, il modello componenziale proposto e descritto dallo psicologo Scherer in numerosi suoi lavori (Scherer, 1984, 1986, 2003; Banse & Scherer, 1996). Per conciliare la contraddizione tra le teorie secondo cui i movimenti muscolari delle espressioni facciali generati da bisogni di segnalazione comunicativa indicano un processamento di informazione e tendenze all'azione, e la ben più diffusa convinzione che le espressioni facciali servano prima di tutto alla trasmissione delle emozioni, Scherer propone quello che egli definisce *Component Process Model of Emotion* (CPM). Il modello componenziale di Scherer scaturisce da una serie di considerazioni derivate da studi sulle espressioni facciali (tra cui Darwin e Duchenne), secondo le quali le particolari innervazioni che costituiscono la muscolatura del volto umano rispondano ad una duplice funzione in quanto indicano, da un lato, un processamento continuo dell'informazione (*information processing*) che si manifesta con il coinvolgimento della muscolatura facciale nella regolazione degli input sensoriali, dall'altro, tendenze comportamentali adattive che si manifestano con il risultato della valutazione di un evento o di una situazione. Secondo tale modello le espressioni facciali prototipiche delle cosiddette emozioni primarie o fondamentali¹⁰⁶ vengono intese come il risultato di un processo sequenziale cumulativo. Viene perciò ipotizzato che, all'interno di questo processo, una serie di "controlli valutativi dello stimolo" (*stimulus evaluation checks* o *SEC*) innescano o attivano specifici movimenti muscolari facciali considerati adattivi con riferimento al processamento

¹⁰⁴ Cfr. Ortony, A., Clore, G. L., Collins, A. (1988), *The Cognitive Structure of Emotion*, UK: Cambridge University Press.

¹⁰⁵ Cfr. Schröder, M. (2004), *Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis*, PhD thesis, PHONUS 7, Research Report of the Institute of Phonetics, Saarland University, pag. 13.

¹⁰⁶ Vedi a tal proposito Ekman (1972); Izard (1971, 1977); Tomkins (1962, 1963, 1984).

dell'informazione, alla tendenza comportamentale appropriata e all'intenzione comunicativa.¹⁰⁷ Ciascuno di questi *patterns* parziali di risposta facciale (che si susseguono rapidamente per effetto degli *stimulus evaluation checks*) si aggiunge alla precedente espressione portando ad una espressione finale che consiste di un accumulo dei *patterns* dell'innervazione muscolare facciale. L'espressione facciale legata al processamento dell'informazione (*information processing*) consisterebbe in un incremento della tensione dei muscoli della regione degli occhi e delle sopracciglia (allo scopo di potenziare l'attività visiva) o all'attivazione di muscoli che influenzano gli *input* sensoriali della bocca e del naso (nei casi in cui viene esternato uno stimolo spiacevole). Molti dei ricercatori che supportano la continuità filogenetica dei movimenti muscolari facciali al servizio dell'adattamento comportamentale, sottolineano il fatto che la presenza e le caratteristiche di tali movimenti facciali traggano solo parzialmente spiegazione dal loro significato funzionale. Questi sono, infatti, determinati anche dal loro ruolo come segnali espressivi nella comunicazione di tendenze a reazioni e azioni da parte di un organismo che reagisce ad un determinato evento. Come già accennato, il modello componenziale dell'emozione si basa su diversi tipi di valutazione di un evento:

1. verifica della novità (*novelty check*);
2. verifica della piacevolezza (*intrinsic pleasantness check*);
3. verifica della significatività dello scopo (*goal/need significance check*);
4. verifica della possibilità di controllo (*coping potential check*);
5. verifica della compatibilità con la norma (*norm/self-compatibility check*).

Quanto ipotizzato da Scherer con il *Component Process Model of Emotion* per la muscolatura facciale, vale anche per gli organi e le strutture deputate alla produzione vocale (muscoli, cartilagini e superfici mucose che contribuiscono ai processi di respirazione, articolazione e fonazione)¹⁰⁸ che vengono affetti e condizionati in modo sequenziale e cumulativo sulla base delle risposte provenienti dagli *stimulus evaluation checks*. Analogamente a quanto accade per le espressioni facciali, anche gli organi coinvolti nella produzione vocale vengono continuamente influenzati dal flusso di processi cognitivi ed emotivi. Come per il volto, l'espressione vocale delle emozioni prototipiche o fondamentali¹⁰⁹ viene vista come un risultato sequenziale, un processo cumulativo di cambiamenti nella produzione della voce che confluiscono in una configurazione prototipica della

¹⁰⁷ Cfr. Scherer, K. R. (1988), *Criteria for Emotion-Antecedent Appraisal: A Review*, in: *Cognitive Perspectives on Emotion and Motivation*, (V. Hamilton, G. H. Bower and N. H. Frijda, eds.), Dordrecht: Kluwer, pp. 89-126.

¹⁰⁸ Vedi a tal proposito i cenni preliminari forniti nel paragrafo 2.2.

¹⁰⁹ Che in questo caso Scherer (1984) definisce "*modal*" emotions.

vocalizzazione in una finestra temporale di un episodio emotivo. A ciascuna di queste componenti corrisponde infatti un elenco di caratteristiche vocali che, in base all'elaborazione del modello cognitivo di una emozione o famiglia di emozioni, permette automaticamente di predirne anche le caratteristiche acustiche.¹¹⁰

Data la sua rilevanza nell'ambito che qui viene trattato, si ritornerà su questo aspetto in maniera più approfondita e dettagliata nel prossimo paragrafo.

¹¹⁰ Cfr. Poggi, I., Magno Caldognetto, E. (2004).

4 Emozioni e voce

Come più sopra accennato, la produzione della voce nell'uomo, e quindi dei suoni del linguaggio, avviene attraverso l'adattamento specifico di una serie di organi deputati ad altre funzioni fisiologiche come la respirazione, la deglutizione ecc. La produzione e trasmissione delle emozioni per mezzo della voce si realizza grazie alla variazione di specifici *pattern* acustici che sono determinati da cambiamenti nella respirazione, fonazione e articolazione concomitanti o conseguenti all'evento emotivo.

Secondo Johnstone & Scherer (2000) il contesto dell'evoluzione darwinianamente intesa avrebbe selezionato e favorito per alcune emozioni maggiori caratterizzazioni nella sfera visiva, mentre per altre nella sfera acustica: per la rabbia e la paura sarebbe stata maggiormente sviluppata l'espressione vocale, in modo che gli antenati dell'uomo potessero avvertirsi e minacciarsi in modo esplicito, anche a lunghe distanze; per emozioni quali disgusto e gioia sarebbero invece stati selezionati e favoriti sviluppi biologici relativi alla mimica facciale, di modo che fra individui di uno stesso gruppo ci si potesse comprendere al volo. Si tratta di una sostanziale differenza tra espressioni facciali e vocali che è strettamente connessa alla trasmissione delle caratteristiche dei segnali utilizzati e che determinano la loro utilità in differenti contesti comunicativi. Mantenendo la terminologia più sopra utilizzata, nel caso delle espressioni facciali il ricevente deve essere necessariamente posto direttamente di fronte all'emittente e ad una distanza tale che gli consenta di differenziare i *patterns* che costituiscono le complesse espressioni facciali: da questo punto di vista l'espressione facciale rappresenta un sistema di comunicazione direzionale e di breve distanza. Quello delle espressioni vocali è, al contrario, un sistema di comunicazione omnidirezionale: percepiamo un suono anche quando non siamo concentrati su di esso e lo percepiamo anche a grandi distanze. Data questa sua peculiarità il canale vocale rappresenta un validissimo sistema di comunicazione di tipo non direzionale e di elevata portata che lo rende idoneo alla trasmissione di situazioni di pericoli (grida di paura o urla di allarme) e alla trasmissione di segnali legati al dominio del territorio che comportano una minaccia ad un numero indefinito di potenziali contendenti.

Che il canale vocale sia uno dei canali preferenziali per l'espressione di stati emotivi, ma non l'unico, emerge anche da uno studio condotto da Albert Mehrabian, che nel volume *Silent messages* (1981), affronta una serie di aspetti della comunicazione non verbale, ivi inclusi movimenti e postura del corpo, espressioni facciali, qualità e intonazione della comunicazione orale, intensità e velocità di eloquio, sottili variazioni nella formulazione delle frasi (che rivelano significati nascosti in ciò che viene detto), come anche questioni riguardanti la combinazione di messaggi e informazioni provenienti da sorgenti diverse (ad esempio espressioni facciali, tono di voce, parole). È in questo volume che, come preannunciato nell'introduzione, l'autore riporta una dettagliata analisi di alcune sue ricerche riguardo messaggi incoerenti relativi a sentimenti e attitudini di tipo *like-dislike* (gradevole-non gradevole) e alle relazioni che intercorrono tra parole e caratteristiche non verbali veicolate attraverso tali messaggi. Da queste ricerche all'autore è stato possibile definire e sintetizzare il grado di importanza dei tre elementi basilari della comunicazione *faccia-a-faccia* (parole, voce ed espressioni facciali) nella trasmissione di messaggi di questo tipo in una sorta di equazione che assegna al contenuto verbale un peso del 7%, al contenuto vocale un peso del 38%, mentre all'espressione facciale viene assegnato un 55% (7%-38%-55%).¹¹¹ L'equazione proposta dallo psicologo (conosciuta anche come *3 Vs rule* - "verbal, vocal, visual") indica che un ascoltatore, o colui che riceve da parte di un altro soggetto un messaggio relativo al proprio stato d'animo o alle proprie attitudini, per decodificare il messaggio farebbe ricorso per il 7% al contenuto verbale (le parole), per il 38% alle caratteristiche vocali (tono di voce ecc.) e per il restante 55% al linguaggio del corpo (espressioni facciali ecc.). Nella comunicazione di messaggi ambigui come quelli a cui l'autore si riferisce, risulta perciò evidente come un ascoltatore, si affidi nel complesso all'informazione veicolata da caratteristiche di tipo non verbale (38% + 55%), piuttosto che all'informazione veicolata dal contenuto verbale.

Ciononostante, sebbene le conclusioni a cui è giunto di Mehrabian riguardino due specifiche situazioni sperimentali (*cfr.* nota 111), la sua equazione è stata spesso fraintesa, portando molti ad estendere l'equazione alla comunicazione nel suo complesso, facendo erroneamente ritenere che in qualsiasi contesto comunicativo il significato di un messaggio sia veicolato in gran parte da caratteristiche non verbali, piuttosto che dal significato delle parole.

¹¹¹ *Cfr.* Mehrabian, A., Wiener, M. (1967), Decoding of inconsistent communications, *Journal of personality and social psychology* 6(1): 109-114 e Mehrabian, A., Ferris, S. R. (1967), Inference of attitudes from nonverbal communication in two channels, *Journal of consulting psychology* 31(3): 248-252.

Nell'atto del parlare, e quindi nella produzione della voce, le variazioni vocali legate all'intonazione e alle pause portano con sé una grande varietà di funzioni sintattiche e pragmatiche e, in alcune lingue, anche semantiche. Analogamente, nel caso delle espressioni facciali sono, invece, le varie parti superiori del volto ad assolvere a funzioni paralinguistiche e conversazionali molto importanti trasmesse dall'emittente al ricevente.¹¹²

Risulta pertanto evidente come i fattori coinvolti, nonostante il breve accenno, siano di numero piuttosto elevato e che nel loro insieme hanno reso e rendono difficile lo studio delle emozioni espresse per mezzo della voce. Chiusa questa brevissima parentesi introduttiva, in questo capitolo si analizzeranno perciò i rapporti che intercorrono tra voce ed emozioni anche alla luce di quanto esposto nei precedenti capitoli, per comprendere più a fondo quanto accade quando voce ed emozioni si fondono in un unico elemento comunicativo. È perciò a quel 38% che compone l'equazione di Mehrabian che ci si rivolgerà in questa sede.

4.1 Informazioni trasmesse dalla voce

Quello del parlato emotivo è un dominio in cui vengono influenzati contemporaneamente più livelli di informazione che Magno Caldognetto *et al.* (2008) hanno raggruppato e definito in termini di:

- informazioni extralinguistiche;
- informazioni paralinguistiche;
- informazioni linguistiche.

Tutte e tre hanno carattere informativo, ma solo le ultime due risultano codificate ed assolvono a funzioni comunicative.¹¹³

Come è facile immaginare si tratta di livelli di informazione che si realizzano in contemporanea e dove ciascun livello di informazione modifica gli altri nello stesso istante in cui viene modificato.

Un esempio di quanto sia difficile separare i livelli di informazione viene da una schematizzazione fornita da Magno Caldognetto (2002), e riprodotta in Figura 27, in cui viene

¹¹² Nel primo caso si veda ad esempio Cutler, A., Ladd, D. R. (Eds.) (1983), *Prosody: Models and measurements*, Berlin, Springer; nel secondo caso si veda invece Ekman, P. (1979), About brows: Emotional and conversational signals, in *Human ethology*, (J. Aschoff, M. von Carnach, K. Foppa, W. Lepenies, D. Plog, Eds.), Cambridge: Cambridge University Press, pp. 169-202.

¹¹³ Occorre fare qui una piccola precisazione sull'uso dei termini 'informativo' e 'comunicativo' con riferimento ai segnali utilizzati nel parlato. Citando Laver (1994: 21): "A signal is communicative, if 'it is intended by the sender to make the receiver aware (to inform him) of something of which he was not previously aware'. A signal is informative, if regardless of the intentions of the sender, 'it makes the receiver aware of something of which he was not previously aware'."

evidenziata la molteplicità di informazioni trasmesse, nel caso specifico, dalla sola frequenza fondamentale (F_0).

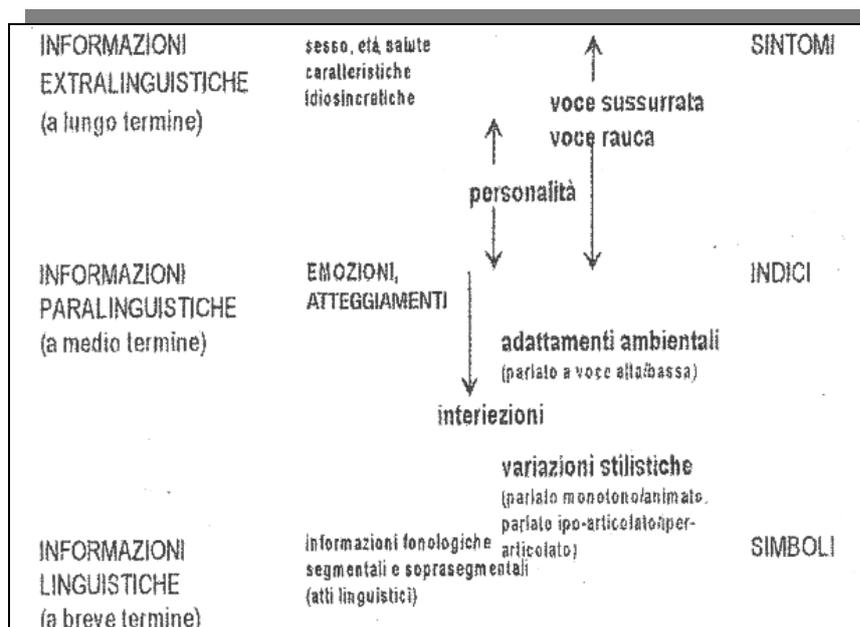


Figura 27: Rappresentazione schematica delle informazioni veicolate dalla frequenza fondamentale (F_0)¹¹⁴

A tal riguardo Magno Caldognetto *et al.* (2008: 187) fanno osservare come, nonostante le difficoltà a cui si è appena accennato e sui cui si ritornerà più approfonditamente, il ricorso alle interiezioni primarie (Poggi, 1981)¹¹⁵ consenta di ridurre al minimo le interazioni tra quelle che sono le caratteristiche acustiche del messaggio verbale e le caratteristiche vocali delle emozioni: in queste produzioni, infatti, i profili intonativi e prosodici paralinguistici si presentano in un contesto più semplificato escludendo, o quanto meno riducendo, l'interazione tra caratteristiche fonologiche e suprasegmentali del parlato emotivo.

Nonostante queste difficoltà si tenterà di analizzare, per quanto possibile, i vari livelli di informazione singolarmente.

4.1.1 Informazioni extralinguistiche

Rientrano nelle informazioni di tipo extralinguistico tutte quelle informazioni o caratteristiche legate alla qualità della voce del soggetto e che sono date dalle caratteristiche anatomo-fisiologiche dell'apparato fonatorio, dalla configurazione delle corde vocali, dalla

¹¹⁴ Fonte: Magno Caldognetto (2002), pag. 203.

¹¹⁵ Si tratta di quelle che Scherer (1994: 170) definisce *affect bursts*: “[...] very brief, discrete, nonverbal expressions of affect both in face and voice as triggered by clearly identifiable events”. Uno degli esempi riportato dallo stesso autore per spiegare cosa egli intenda per *affect burst* è: “a facial/vocal disgust expression upon seeing a hairy black worm emerging from an oyster shell one is about to bring closer to one’s mouth”.

loro tensione e dalla loro frequenza di vibrazione. Si tratta, quindi, di “caratteristiche anatomiche permanenti ed esclusive dell’individuo”.¹¹⁶

Riguardo alle qualità della voce del soggetto queste sono determinate dalla conformazione fisiologica delle corde vocali (spessore, lunghezza, elasticità) che, con la loro vibrazione, che più sopra è stata definita frequenza di vibrazione delle corde vocali (*frequenza fondamentale* o F_0), danno origine a quelli che possono essere definiti elementi caratterizzanti di ciascun individuo.¹¹⁷

Si parlerà, in questo caso, di elementi caratterizzanti, ovvero di caratteristiche che variano in funzione di:

- sesso: la conformazione delle corde vocali è legata a quella del sesso come anche alle dimensioni e alla forma del condotto laringale che porta ad avere una F_0 più alta nelle donne (dovuta a corde più tese e sottili e condotto laringale più ristretto) rispetto agli uomini; ulteriori differenze si hanno anche nei bambini rispetto agli adulti, dove i primi, per le stesse ragioni, hanno una F_0 ancora più alta;¹¹⁸
- stato di salute: frequenza e modalità di vibrazione delle corde vocali sono determinate dallo stato di salute del parlante che, a seconda delle patologie, produce voce rauca, laringalizzata, soffiata ecc. Le cause sono attribuibili sia alla presenza di patologie come malformazioni alle corde (noduli, polipi ecc.) che ne impediscono una corretta vibrazione, come anche a cause indirette quali quelle indotte da fumo, droghe, alcool che ne alterano le caratteristiche originarie: ne è un esempio la voce dei soggetti fumatori che al nostro orecchio suona come rauca per l’ispessimento e l’accumulo di mucosa sulle corde vocali;¹¹⁹
- altezza e peso: le caratteristiche extralinguistiche sono strettamente correlate alla corporatura del soggetto che ne influenza condotto vocale, corde vocale e organi articolatori deputati alla produzione dei suoni del linguaggio. A parità di suono prodotto e quindi articolato, secondo la teoria proposta da Fant (1960) si hanno differenze a livello spettrale legate alle caratteristiche di risonanza delle strutture coinvolte.

¹¹⁶ Cfr. Anolli, L. (2006), *Fondamenti di psicologia della comunicazione*, Bologna, Il Mulino.

¹¹⁷ Vedi Laver, J. (1994), *Principles of Phonetics*, Cambridge: Cambridge University Press.

¹¹⁸ Riguardo questo aspetto si veda quanto già riportato in nota 27.

¹¹⁹ Per una trattazione più approfondita e dettagliata si veda Verdolini, K. (1994), *Voice disorders*, in: J. Bruce Tomblin, H. L. Morris, D. C. Spriestersbach (Eds.), *Diagnosis in speech-language pathology*, San Diego: Singular Publishing Group, pp. 247-306.

Si assiste, in poche parole, ad una elevata variabilità dovuta anche ad aspetti idiosincratici legati al parlante dove la possibilità di isolare determinate caratteristiche diventa estremamente complessa e in alcuni casi addirittura impossibile a causa di elementi concomitanti che nel prosieguo saranno esaminati.¹²⁰

Si tratta, quindi, di caratteristiche che non hanno, come già detto, carattere comunicativo, ma possono, invece, essere considerate di tipo esclusivamente informativo circa l'individuo.

4.1.2 Informazioni paralinguistiche

Le informazioni di tipo paralinguistico rappresentano quell'insieme di proprietà acustiche transitorie che accompagnano la pronuncia di un enunciato e che possono variare da situazione a situazione con funzioni sia informative che comunicative.

È sulla base di tali proprietà che, ai vari ricercatori che se ne sono occupati, è stato possibile individuare determinate caratteristiche acustiche che veicolano l'espressione del parlato emotivo, come conferma anche quanto formulato da Canepari (1985: 87):

“Intendiamo paralinguistica l'insieme di quegli elementi fonici non-verbali che rendono completa e organica la comunicazione verbale umana, liberandola dall'artificiosità e dalla meccanicità cui sarebbe condannata se si potesse/dovesse servire solo degli elementi linguistici in senso stretto.”

Per l'individuazione delle variazioni di tali proprietà acustiche nel parlato emotivo vengono di norma eseguite misurazioni nel dominio della frequenza, dell'intensità e del tempo.¹²¹

Riguardo al dominio della frequenza vengono ricavati per F_0 (*pitch*) i valori massimi e minimi, il *range*¹²², il valore medio, la mediana e relativa deviazione standard, compreso lo *slope* (dato dal rapporto tra *range* di F_0 e durata dell'enunciato).

Riguardo al dominio dell'intensità vengono calcolati i valori massimi, minimi, *range* globale, le concentrazioni di energia in diverse bande spettrali.

¹²⁰ Ne sono un esempio i casi giudiziari dove la possibilità di identificare un individuo sulla base della voce risulta fortemente limitata da distorsioni e disturbi provenienti dall'ambiente circostante e da caratteristiche legate ai canali di comunicazione attraverso i quali la voce viene captata. Per una rassegna di tali problematiche e delle metodologie comunemente utilizzate in questo ambito si veda ad es. Hollien, H. (2002), *Forensic voice identification*, Academic Press.

¹²¹ In Appendice I viene riprodotta una tabella tratta da Scherer, Johnstone, Klassmeyer (2003) contenente i parametri acustici e psicoacustici più comunemente utilizzati nelle ricerche sul parlato emotivo.

¹²² Con *range* viene indicata l'estensione della gamma dei valori di F_0 .

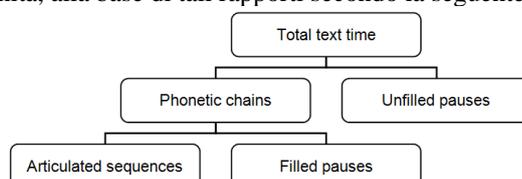
Per il dominio del tempo vengono invece misurate le durate (assolute e relative) di foni, di sillabe, di parole o dell'intero enunciato, da cui, infine, vengono ricavati i vari rapporti riconducibili a:

- *articulation rate*: numero di sillabe diviso per il tempo della sequenza articolata (syll./s);
- *speech rate*: numero di sillabe diviso per il tempo della catena fonica (syll./s);
- *fluency rate*: numero di sillabe diviso per la durata del testo totale;
- *word rate*: numero di parole diviso per il tempo della sequenza articolata (words/s).¹²³

Da uno studio condotto da Breitenstein, Van Lancker, Daum (2001) lo *speech rate* si è rivelato essere un parametro utilizzato dagli ascoltatori per differenziare alcune emozioni veicolate attraverso la voce: utilizzando stimoli vocali emotive modificandone sistematicamente *pitch* e durata hanno dimostrato che gli ascoltatori associavano uno *slow rate* con le etichette di *sadness*, mentre un *fast rate* veniva, invece, associato ad etichette emotive come *angry*, *frightened* o *neutral*. Un altro studio condotto da Kienast, Paeschke & Sendlmeier (1999) suggerisce, infatti, come l'accuratezza degli aspetti articolatori e di durata possano essere considerati parametri utili a caratterizzare, se non proprio tutte, almeno alcune emozioni (come ad esempio *anger*, *fear* e *sadness*).

Ai parametri appena elencati si aggiungono anche *jitter* e *shimmer* che, come già detto, forniscono informazioni su comportamenti e caratteristiche a-modalità della voce. Questi

¹²³ Sul calcolo di tali indici sussistono, però, una serie di problematiche connesse, sostanzialmente, alla definizione degli elementi (sillaba, catena fonica, pausa piena, pausa vuota ecc.) utilizzati per il calcolo dei vari rapporti. Per semplicità si farà qui riferimento a Magno Caldognetto, Zmarich, Ferrero (1997: 779) che definiscono gli elementi, o le unità, alla base di tali rapporti secondo la seguente organizzazione gerarchica:



e con le seguenti definizioni:

- *total text*: the whole speech production, i.e. the sum of speech chains, unfilled pauses, filled pauses, disfluencies;
- *unfilled pauses*: any silent or breathing interval between two successive phonetic chains, excluding the closure phase of any voiceless plosive consonant starting the successive phonetic chain;
- *phonetic chains*: the sequence of phonetic segments (including filled pauses and disfluencies) delimited by two silent pauses;
- *filled pauses*: any occurrence of hesitations, interjections, abnormally lengthened vowels, repetitions etc.;
- *articulated sequence*: any phonetic chains excluding all filled pauses”.

Per una rassegna più ampia delle varie definizioni presenti in letteratura si veda anche Romito, L., Galatà, V., Lio, R. (2006), Fluency Articulation and Speech Rate as new parameters in the Speaker Recognition, in: *Atti del convegno "III Congreso de Fonética Experimental"*, Santiago de Compostela, 26-24 ottobre, 2005, Xunta de Galicia, pp. 537-549.

parametri sono strettamente connesse al modo di vibrare delle *pliche* vocali e sono all'origine dei vari tipi di fonazione (come ad es. *breathy*, *whispery*, *creaky* e *harsh*) che ne determinano anche la qualità (Laver, 1980). In questo caso, ad esempio, il disgusto viene caratterizzato rispetto alla tristezza dalla presenza di valori elevati di *shimmer* e *jitter*.¹²⁴

In alcuni casi (Banse & Scherer, 1996; Scherer, 2000; Chasaide & Gobl, 1997) si è fatto ricorso a studi sulle qualità non modali della voce, soprattutto perché non tutte le emozioni vengono caratterizzate in modo univoco in base ai parametri di cui si è appena discusso e che vengono considerate anche caratteristiche macroprosodiche.¹²⁵

Altre misurazioni possibili sono rappresentate dal calcolo del rapporto tra armoniche e rumore (*Harmonics to Noise Ratio*, HNR) presente nel segnale, dall'indice di Hammarberg o ancora dal calcolo dei valori formantici delle vocali con le rispettive larghezze di banda.¹²⁶

L'analisi dei parametri sin qui descritti viene di norma condotta su catene (*chunk*) di parlato emotivo di lunghezza variabile (si va da foni, sillabe e parole a frasi o interi brani).

Per la valutazione delle misurazioni effettuate nei vari domini di analisi si fa, in genere, ricorso a particolari processi di normalizzazione che, solitamente, come ad esempio in Anolli *et al.* (2008a, 2008b) prevedono l'assunzione di una *baseline* o di un valore di riferimento al parlato non emotivo (ovvero alla produzione neutra).¹²⁷ In base a processi di normalizzazione che consentono di esprimere e quantificare la distanza dei singoli parametri dai parametri di riferimento del parlato non emotivo (di norma in termini di valori medi con analisi di varianza, correlazione o clustering) è stato inoltre possibile stabilire quella che

¹²⁴ Cfr. Poggi, Magno Caldognetto (2004). Tuttavia, le variazioni legate ai valori di *shimmer* sono meno significative nella produzione e nella percezione delle emozioni rispetto a quelle connesse ai valori di *jitter* (vedi Klingholz & Martin, 1985).

¹²⁵ Cfr. Magno Caldognetto *et al.* (2008: 191), in cui il concetto di macroprosodico viene contrapposto a quello di microprosodico, ovvero un livello di analisi che consente di mettere in evidenza e di localizzare specifiche modificazioni imposte dall'espressione delle emozioni rispetto a quelle che possono essere considerate le caratteristiche prosodiche e intonative richieste dalle diverse tipologie frasali nelle varie lingue. Attraverso un'analisi di tipo microprosodico è possibile individuare la collocazione dei picchi di F_0 e di intensità rispetto alle vocali toniche presenti nel *continuum*, o ancora all'andamento di F_0 sulle sillabe finali di un sintagma o di un enunciato. Magno Caldognetto *et al.* (*op. cit.*) fanno rilevare come un'analisi di tipo macroprosodico permetterebbe di meglio valutare due ipotesi alternative proposte, rispettivamente da Scherer *et al.* (1984) e Ladd *et al.* (1985), riguardo la produzione delle caratteristiche prosodiche del parlato emotivo: la teoria dei *canali paralleli* (in cui le informazioni linguistiche e paralinguistiche vengono codificate indipendentemente l'una dall'altra per poi fondersi in una fase successiva) e la teoria del *canale unico* (che contempla, invece, un'unica pianificazione causata e determinata da uno specifico atto linguistico). Per un'analisi in questi termini si veda Kori, Magno Caldognetto (2003) in cui viene operato un confronto tra italiano e giapponese.

¹²⁶ Per maggiori dettagli si veda Magno Caldognetto (2002). Rispetto a questo parametro, ad esempio, un elevato HNR caratterizza la tristezza rispetto alle altre emozioni. In altri casi si è fatto ricorso anche allo spettro a lungo termine della voce (*Long Term Spectrum*, LTS) definito come spettro di intensità media calcolato per specifiche gamme di frequenza sul parlato continuo.

¹²⁷ Altri processi di normalizzazione, in questo caso per F_0 , prevedono la conversione da misure acustiche (espresse in Hertz) a misure uditive (espresse in ottave e semitoni), come ad esempio in Kori & Magno Caldognetto (2003), dove la conversione dei valori di F_0 è stata effettuata sulla base dei rispettivi valori in semitoni secondo la formula $12 * \log(F_0 \text{ in Hz}) / \log(50)$. Vedi anche Ladd & Terken (1995) e Johnson (2005).

Magno Caldognetto *et al.* (2008: 188) definiscono “gerarchia di rilevanza”, ovvero un ordine gerarchico in cui alcuni elementi acustici influiscono più degli altri nell’espressione delle emozioni vocali stabilendo livelli di significatività e gradi di dipendenza tra gli indici analizzati. In base a questo ordine gerarchico si hanno, per rilevanza, dapprima informazioni legate a F_0 a cui seguono, nell’ordine, informazioni sull’intensità e informazioni sulla durata.

A fronte della diversità tra le lingue e i tra i singoli parlanti, sulla base delle predette analisi è stato possibile isolare per le emozioni primarie delle caratteristiche acustiche sufficientemente stabili, nonostante qualche lieve differenza riconducibile a diverse tipologie di parlanti o ancora alla diversità del materiale linguistico analizzato (Pittam & Scherer, 1993; Johnstone & Scherer, 2000; Scherer, 2003; per l’italiano Kori & Magno Caldognetto, 1986; Magno Caldognetto *et al.*, 1998).

Se si esaminano, infatti, i risultati di alcuni studi passati in rassegna da Johnstone & Scherer (2000) che si basano sui parametri appena presentati, per le quattro emozioni più studiate (*anger*, *fear*, *sadness* e *joy*) si rileva, ad esempio, come tali parametri siano assai significativi nella caratterizzazione delle emozioni vocali. Per l’emozione vocale di *anger* si ha in generale un incremento di F_0 media e dell’intensità media. I contorni di F_0 tendono verso il basso mentre lo *speech rate* tende ad aumentare nel caso di *hot anger*.¹²⁸ Per l’emozione di *fear* (caratterizzata da livelli alti di *arousal*) è stato riportato un incremento dell’intensità di F_0 media e di F_0 range, anche se questo dato sia discordante con Banse & Scherer (1996) che trovarono, invece, un sostanziale decremento nel range di F_0 . Lo *speech rate* viene riportato come elevato. Anche in realizzazioni caratterizzate da basso *arousal*, sono stati rilevati articolazioni più rapide, mentre gli altri risultati appaiono meno coerenti risultanti sia in un incremento che in un decremento di F_0 media ed F_0 range. Per la tristezza sono stati invece registrati abbassamenti di F_0 media, F_0 range e intensità (sebbene Banse & Scherer, 1996, riportino un incremento del range di F_0), come anche una tendenza verso il basso dei contorni di F_0 . Anche lo *speech rate* si abbassa nel caso della tristezza. Per la gioia si riscontrano, invece, alcune evidenti incongruenze legate al livello di *arousal* dell’emozione studiata: nel caso di gioia caratterizzata da elevato *arousal* si rileva, ad esempio, un incremento di F_0 media, del livello di F_0 e dell’intensità come anche del range di F_0 .

In Amir *et al.* (2003), ad esempio, i contorni intonativi per alcune frasi *angry* sono stati alterati manualmente: dopo aver presentato gli stimoli a gruppi di soggetti ascoltatori,

¹²⁸ Johnstone & Scherer (2000) riportano altresì come in alcuni studi si possano trovare anche valori elevati con riferimento alla variabilità e al range di F_0 , dovuti probabilmente alla mancata distinzione tra *hot anger* e *cold anger*, mentre in altri, in cui non sono stati riscontrati valori simili, sono state probabilmente analizzate produzioni di *cold anger*.

questi percepivano le frasi come meno *angry*, confermando come le informazioni trasmesse dal *pitch* rappresentino un essenziale indice per il riconoscimento di stati emotivi nel parlato.

	F ₀ media	F ₀ range	F ₀ variabilità	F ₀ contorno	Intensità media	Intensità > 1kHz	Speech rate
Collera	+	++	++	disc.	++	++	+
Paura	++	++	++		+	++	++
Tristezza	-	-	-	disc.	--	-	-
Gioia	+	+	+		+		+
Disgusto	+	(-)	(-)				-

Tabella 4: Schematizzazione delle caratteristiche acustiche delle emozioni primarie¹²⁹

Da quanto appena riportato, e sintetizzato nella Tabella 4, si evince come per alcune emozioni vi sia una condivisione di indici acustici, mentre si distinguono nettamente per altri. Come evidenzia Scherer (2003), la coerenza e la stabilità di tali caratteristiche non è dovuta soltanto all'universalità nell'espressione delle emozioni primarie, quanto anche e soprattutto al metodo di raccolta dei dati dove, per i motivi a cui si accennerà nel capitolo 6, la produzione del parlato emotivo viene nella maggior parte dei casi affidata ad attori sulla base di etichette linguistiche e scenari che producono forme che possono, in alcuni casi, sollevare critiche riguardo il loro essere stereotipate e prototipiche.

4.1.3 Informazioni linguistiche

Le informazioni linguistiche sono rappresentate da informazioni di tipo segmentale e suprasegmentale, da informazioni di tipo fonologico e grammaticale, ovvero di tutte quelle informazioni relative a vocali e consonanti, ivi comprese la loro specifica conformazione spettrale, lo stato vibratorio periodico e a-periodico delle pliche vocali, le transizioni spettrali e gli effetti di fenomeni di coarticolazione indotti dagli organi fono-articolatori nel loro movimento e passaggio da una posizione all'altra nella produzione dei singoli suoni. Ne fanno parte, naturalmente, anche tutte quelle informazioni di intensità, variazioni di F₀, durata di singoli foni, pause e di quant'altro di dipendente dalla struttura sintattica e pragmatica degli enunciati. Si tratta di un dominio di informazione che può essere definito omnicomprensivo, un livello di informazione in cui vengono coinvolti sia i meccanismi di produzione, sia i meccanismi di percezione del segnale acustico/verbale nel suo complesso.¹³⁰

¹²⁹ Tabella tratta da Magno Caldognetto *et al.* (2008: 188) elaborata sulla base dei risultati di Pittam & Scherer (1993) modificata secondo Johnstone & Scherer (2000).

¹³⁰ Si tratta di un ambito assai vasto la cui trattazione in termini esaustivi si rende qui particolarmente complicata. Accanto a quanto già in parte detto nel capitolo 2 si faccia riferimento ai più comuni manuali di Fonetica e *Speech Communication* oggi disponibili.

4.2 Le emozioni vocali nel Component Process Model di Scherer

Nell'affrontare la questione delle emozioni vocali nel *Component Process Model* di Scherer non mancheranno riferimenti, con parallelismi e differenze, alle espressioni facciali che serviranno a meglio comprendere alcuni concetti chiave attorno alle quali si sono sviluppate le indagini sul parlato emotivo.

Come precedentemente accennato, Scherer (1994) sostiene che, analogamente all'espressione facciale, anche l'espressione vocale sia filo-geneticamente continua e che l'origine di molte forme espressive vocali (soprattutto di quelle che egli definisce *affect bursts*) sia da ricondurre a significati di tipo funzionale quali *input* sensoriali (chiusura o restringimento della cavità orale o nasale), intenzioni comportamentali (espulsioni o tentativi di vocalizzazione) e segnali comunicativi (minaccia, ecc.). Sebbene ci si rivolgerà in questa sede solo alle espressioni vocali delle emozioni, un raffronto, seppur parziale, con le espressioni facciali, ci consentirà di meglio definire l'argomento qui trattato.

Secondo quanto teorizzato da Scherer nel suo *Component Process Model of Emotions*, sembrerebbe che le espressioni facciali e le espressioni vocali siano organizzate biologicamente e culturalmente in maniera parallela.

	Neutral	Anger	Fear	Joy	Sadness	Disgust	Surprise	Mean
Facial/Western/20	-	78	77	95	79	80	88	78
Vocal/Recent Western/11	74	77	61	57	71	31	-	62
Facial/Non-Western/11	-	59	62	88	74	67	77	65
Vocal/Non-Western/1	70	64	38	28	58	-	-	52

Tabella 5: Accuratezza (%) nel riconoscimento di emozioni facciali e vocali in paesi occidentali e non-occidentali¹³¹

Se così fosse, secondo Scherer ci sarebbe da chiedersi fino a che punto questi due sistemi di comunicazione e segnalazione siano ridondanti o fino a che punto i due sistemi possano, invece, sovrapporsi per trasmettere la stessa informazione a degli osservatori esterni. Come emergerà più avanti, e come parzialmente desumibile dalla Tabella 5, esistono non solo analogie tra i due sistemi espressivi, ma anche sostanziali differenze.

I continui cambiamenti delle innervazioni muscolari del volto non producono sempre differenze o modificazioni visibili sulla pelle del volto e i cambiamenti o le variazioni degli organi coinvolti nella produzione vocale non sono sempre udibili. C'è tuttavia una differenza ancora maggiore. Potrebbe, infatti, sembrare che la visibilità dei cambiamenti nel livello di tensione della muscolatura facciale sia prevalentemente dovuta all'intensità dell'innervazione

¹³¹ Tabella tratta ed adattata da Scherer (2003), pag. 236. I valori riportati in tabella per le espressioni facciali si riferiscono allo studio di Ekman (1994), mentre i valori per le espressioni vocali si riferiscono allo studio di Scherer, Banse, Wallbott (2001).

e all'intenzione comunicativa del esprimente che, di fatto, comporterebbe un'assenza di capacità di controllo o regolazione o di amplificazione di un segnale debole a scopi comunicativi. Di contro, nell'espressione vocale il *processing* dell'informazione e/o dell'emozione produrrebbero cambiamenti nella qualità della voce udibili solo se un suono viene prodotto (ovvero, se si ha vibrazione delle corde vocali con la produzione di impulsi glottali che, una volta filtrati nelle cavità soprastanti, portano ad un segnale acustico che è possibile udire o registrare su un qualsiasi supporto). Soltanto durante la vocalizzazione gli effetti dell'*information processing*, come tendenze respiratorie, fonazione e parametri articolatori verranno esternalizzati e resi disponibili alle indagini e allo scrutinio dell'osservatore. Per questo motivo l'udibilità di determinati *patterns* acustici legati alla voce non è esclusivamente determinata dall'intensità del soggiacente eccitamento causato da un evento emotigeno, ma anche e soprattutto dall'attivazione o disattivazione delle vocalizzazioni innescate da impulsi motori (consapevoli o inconsapevoli). A ciò si aggiunga come, tra l'altro, resti a noi la facoltà di attivare la vocalizzazione quando vogliamo parlare e di disattivarla quando vogliamo tacere: poiché il parlato è una delle attività umane più controllate, la disponibilità di segnali vocali non verbali di *information processing* che accompagnano continuamente la vocalizzazione, è sempre determinata dall'attività verbale consciamente controllata. Le espressioni facciali sono, dunque, sempre disponibili, mentre le espressioni vocali lo sono solo e soltanto se le corde vocali vibrano e questa è una delle caratteristiche più importanti per comprendere non solo la relazione tra espressioni facciali ed espressioni vocali, ma anche le sostanziali differenze che ne complicano le indagini.

Si è già fatto cenno ai meccanismi coinvolti nel modello componenziale dell'emozione e ai diversi tipi di valutazione di un evento, ma si consideri più da vicino come a ciascuna di queste componenti valutative corrisponda un elenco di caratteristiche vocali (vedi Tabella 6).

Poiché lo stato emotivo di una persona influenza continuamente anche la sua respirazione, la sua fonazione e la sua capacità di articolazione, tutta l'attività verbale fornisce un continua chiave di lettura dello stato d'animo del parlante che funge da importante contesto informativo all'ascoltatore.

Component patterning theory predictions for voice changes (adapted from Scherer, 1986)**Novelty check**

<i>Novel</i>	<i>Not novel</i>
Interruption of phonation, sudden inhalation, ingressive (fricative) sound with glottal stop (noise-like spectrum)	No change

Intrinsic pleasantness check

<i>Pleasant</i>	<i>Unpleasant</i>
Faucal and pharyngeal expansion, relaxation of tract walls, vocal tract shortened by mouth corners retracted upward (increase in low frequency energy, F1 falling, slightly broader F1 bandwidth, velopharyngeal nasality, resonances raised) "Wide voice"	Faucal and pharyngeal constriction, tensing of tract walls, vocal tract shortened by mouth corners retracted downward (more high frequency energy, F1 rising, F2 and F3 falling, narrow F1 bandwidth, laryngopharyngeal nasality, resonances raised) "Narrow voice"

Goal/need significance check

<i>Relevant and consistent</i>	<i>Relevant and discrepant</i>
Overall relaxation of vocal apparatus, increased salivation (F0 at lower end of range, low-to-moderate amplitude, balanced resonance with slight decrease in high-frequency energy) "Relaxed voice" If event conducive to goal: relaxed voice + wide voice If event obstructive to goal: relaxed voice + narrow voice	Overall tensing of vocal apparatus, decreased salivation (F0 and amplitude increase, jitter and shimmer, increase in high frequency energy, narrow F1 bandwidth, pronounced formant frequency differences) "Tense voice" If event conducive to goal: tense voice + wide voice If event obstructive to goal: tense voice + narrow voice

Coping potential check

<i>Control</i>	<i>No control</i>
(As for relevant and discrepant)	Hypotonus of vocal apparatus (low F0 and restricted F0 range, low amplitude, weak pulses, very low high-frequency energy, spectral noise, format frequencies tending toward neutral setting, broad F1 bandwidth) "Lax voice"
<i>High power</i>	<i>Low power</i>
Deep, forceful respiration, chest register phonation (low F0, high amplitude, strong energy in entire frequency range) "Full voice"	Head register phonation (raised F0, widely spaced harmonics with relatively low energy) "Thin voice"

Norm/self-compatibility check

<i>Standards surpassed</i>	<i>Standards violated</i>
Wide voice + full voice (+ relaxed voice if expected or + tense voice if unexpected)	Narrow voice + thin voice (+ lax voice if no control or + tense voice if control)

Tabella 6: Elenco delle modificazioni nella qualità della voce attese come risultato delle risposte degli stimulus evaluation checks secondo la Component patterning theory di Scherer¹³²

Dalla Tabella 6 sopra riportata, si desume come gli organi coinvolti nella produzione della parola (che si riammenta non essere altro che un adattamento di organi deputati ad altre

¹³² Tabella tratta e adattata da Scherer (2003), pag. 245.

funzioni fisiologiche come la respirazione, la deglutizione ecc.) subiscano, nel caso specifico delle emozioni, una serie di modificazioni che si riflettono e danno origine a variazioni di *patterns* acustici: si tratta, perciò, di variazioni che sono strettamente connesse e determinate da cambiamenti nella respirazione, fonazione e articolazione concomitanti o conseguenti all'evento emotivo. Al contrario delle espressioni facciali, dove si ha, invece, una diversa origine funzionale dei movimenti muscolari che sono piuttosto brevi e discreti e di tipo prevalentemente adattivo (come ad esempio il corrugamento delle sopracciglia, il digrignare i denti ecc.) riconducibili ad un particolare scopo, nell'espressione vocale molti dei cambiamenti che influenzano l'onda acustica sono legati a *patterns* di eccitamento (*arousal*) fisiologico. In base al modello proposto da Scherer e in base alle modificazioni indotte dagli effetti degli *stimulus evaluation checks* sulla produzione della voce, è possibile predire quelli che sono gli effetti delle emozioni su tali *patterns* e sulle caratteristiche prettamente acustiche della produzione vocale. Si tratta, quindi, di risposte specifiche per ciascuna emozione "modale" per le quali è possibile predirne l'espressione vocale prototipica (vedi Tabella 7), predizioni che sono state empiricamente confermate in tutta una serie di studi successivi.¹³³

	ENJ/ HAP	ELA/ JOY	DISP/ DISG	CON/ SCO	SAD/ DEJ	GRI/ DES	ANX/ WOR	FEAR/ TER	IRR/ COA	RAG/ HOA	BOR/ IND	SHA/ GUI
<i>F0</i>												
Perturbation	<=	>			>	>		>		>		>?
Mean	<✓	>✓	>	<>	<>✓	>✓	>?	>>✓	<>✓	<>	<✓	>?
Range	<=	>			<	<		>>	<	>>		
Variability	<	>			<	>?		>>?	<	>>✓		
Contour	<	>			<	>		>>	<	=		>
Shift regularity	=	<						<		<	>	
<i>Formants</i>												
F1 Mean	<	<	>	>	>	>	>	>	>	>	>	>
F2 Mean		<	<	<	<	<	<	<	<	<	<	<
F1 Bandwidth	>	<>	<<	<	<>	<<	<	<<	<<	<<	<	<
Formant precision		>	>	>	<	>	>	>	>	>		>
<i>Intensity</i>												
Mean	<✓	>✓	>?	>>?	<<✓	>✓		>✓	>✓	>>✓	<>	
Range	<=	>			<			>	>	>		
Variability	<	>			<			>	>	>		
<i>Spectral parameters</i>												
Frequency range	>	>	>	>>	>	>>		>>	>	>	>	>
High-frequency energy	<	<>✓	>	>	<>	>>✓	>?	>>	>>	>>✓	<>	>
Spectral noise					>							
<i>Duration</i>												
Speech rate	<?	>✓			<✓	>		>>✓		>✓		
Transition time	>	<			<	<		<		<		

Note: ANX/WOR: anxiety/worry; BOR/IND: boredom/indifference; CON/SCO: contempt/scom; DISP/DISG: displeasure/disgust; ELA/JOY: elation/joy; ENJ/HAP: enjoyment/happiness; FEAR/TER: fear/terror; GRI/DES: grief/desperation; IRR/COA: irritation/cold anger; RAGE/HOA: rage/hot anger; SAD/DEJ: sadness/dejection; SHA/GUI: shame/guilt; F0: fundamental frequency; F1: first formant; F2: second formant; >: increase; <: decrease. Double symbols indicate increased predicted strength of the change. Two symbols pointing in opposite directions refer to cases in which antecedent voice types exert opposing influence. (✓) prediction supported, (?) prediction contradicted by results in (Banse and Scherer, 1996).

Tabella 7: Cambiamenti attesi per alcuni parametri acustici di alcune emozioni sulla base della predizione delle modificazioni attese nella qualità della voce riportate più sopra in Tabella 6¹³⁴

¹³³ Vedi ad es. Scherer (1989); Scherer, Banse, Walbott, Goldbeck (1991).

¹³⁴ Tabella tratta ed adattata da Scherer (2003), pag. 246.

4.3 Dimensioni psicologiche delle emozioni

Gli indici acustici a cui si è sino ad ora fatto riferimento sono strettamente connessi e condivisi con quelle che vengono identificate come dimensioni psicologiche delle emozioni, ovvero attivazione (*activation, arousal*), valutazione (*valence, evaluation*) e potenza (*power, control*).¹³⁵ Le dimensioni psicologiche non sono da ritenersi veri e propri parametri descrittivi, quanto invece variabili indipendenti che predicono specifiche variabili acustiche e sono di tipo prettamente soggettivo.

La dimensione di attivazione, connessa all'importanza dello scopo, identifica emozioni ad elevata attivazione da quelle a bassa attivazione. Nelle prime si rilevano valori di F_0 media e range di F_0 più alti, dinamica di F_0 maggiore, intensità più alta, sintagmi più lunghi, elevato *articulation rate*, pause brevi. Si distinguono in questo modo le emozioni di sorpresa, gioia, collera calda e paura da tristezza e disgusto.

La dimensione di valutazione è invece connessa al raggiungimento dello scopo, o al contrario alla compromissione di esso. Vengono distinte in base a questa dimensione le emozioni positive da quelle negative in cui si assiste ad abbassamenti repentini di F_0 , aumento dell'intensità, aumento dei picchi massimi di intensità, pause più lunghe. Da queste caratteristiche paura e collera si dovrebbero distinguere da gioia.

La dimensione di potenza è strettamente correlata alla possibilità di controllo delle risorse necessarie a far fronte al raggiungimento dello scopo. Vengono così distinte le emozioni a seconda che comportino dominanza o sottomissione. La dominanza è caratterizzata da F_0 media a frequenze piuttosto basse e intensità elevata, e distinguono rabbia da tristezza.¹³⁶

4.4 Nota conclusiva

L'uomo comunica con gli altri membri della propria specie attraverso quello che in queste pagine è stato definito come sistema vocale, uno strumento che gli consente di produrre suoni organizzati lungo un *continuum fonico* che possono essere percepiti, decodificati e interpretati dagli altri suoi simili. Attraverso la comunicazione verbale e non verbale trasmettiamo ai nostri interlocutori molteplici informazioni, informazioni che Magno

¹³⁵ Schröder (2004) fa rilevare come i nomi utilizzati per queste dimensioni sono stati scelti dai vari studiosi per interpretare in un modo o nell'altro i dati ottenuti dalle proprie indagini e non scaturiscono, pertanto, dai dati stessi. In letteratura si assiste ad una certa arbitrarietà nella nomenclatura utilizzata in questo ambito: ad es. Mehrabian e Russell (1974) le chiamano *pleasure, arousal e dominance*; Osgood *et al.* (1957) usano nomi come *evaluation, activity e potency*, Cowie *et al.* (2001) usano *evaluation, activation e power*.

¹³⁶ Cfr. Magno Caldognetto *et al.* (2008), pp. 189-190.

Caldognetto *et al.* (2008) hanno definito come informazioni di tipo linguistico, para-linguistico ed extra-linguistico, livelli di informazione diversa che si sovrappongono e si influenzano nello stesso istante in cui vengono influenzati per ragioni varie che, per quanto possibile, si è cercato di mettere in evidenza analizzandoli singolarmente.

In modo analogo, nel comunicare con gli altri attraverso la voce e grazie a specifiche variazioni nel canale vocale (che influenzano prevalentemente le caratteristiche acustiche delle espressioni vocali prodotte, con ripercussioni su indici nel dominio della frequenza, dell'intensità e della durata), trasmettiamo una serie di informazioni riguardo alle nostre intenzioni e le nostre attitudini affettive. Si è visto, a questo proposito, come determinati parametri o indici acustici contraddistinguano determinate emozioni *primarie* (o *modali* secondo la terminologia di Scherer) e di come vi sia, allo stesso tempo, una certa condivisione e sovrapposizione di tali indici che ne rendono problematica una caratterizzazione esplicita.

Si è poi tentato di dimostrare come sia possibile, ricorrendo al *Component Process Model of Emotion* proposto da Scherer, predire tali variazioni. Alla base delle emozioni sono presenti una serie di meccanismi di *arousal* che agiscono e influenzano il comportamento e il funzionamento degli organi della parola coinvolti nel processo di fonazione e nell'articolazione dei singoli suoni del parlato (organi che altro non sono che un adattamento di elementi deputati ad altre funzioni primarie).

5 Aspetti ed applicazioni tecnologiche degli studi sulle emozioni

Per dare un'idea di quanto le emozioni vocali e soprattutto gli studi che ad esse si rivolgono siano importanti, si apre qui una piccola parentesi con un brevissimo accenno a quelli che possono essere considerati gli aspetti e le applicazioni tecnologiche degli studi sulle emozioni in generale, e sulle emozioni vocali nello specifico.

Quello che qui verrà brevemente trattato rappresenta un dominio che investe il binomio emozioni e tecnologia, e che può essere definito come il punto di contatto e di fusione di quelle che sono le ricerche empiriche con il mondo della tecnologia e delle applicazioni che rientrano nel campo dell'Intelligenza Artificiale (IA).

Si tratta di un aspetto, quello tecnologico, che trova supporto in uno studio sul riconoscimento cross-culturale delle emozioni in nove paesi condotto da Scherer (2000), in cui l'autore si chiede se sia possibile utilizzare algoritmi simili per produrre variazioni emotive in lingue e culture diverse. Per Scherer il punto sta nel riuscire a stabilire il grado di universalità vs. relatività linguistica e/o culturale degli effetti delle emozioni sulla produzione vocale, una questione di notevole importanza per l'abbattimento dei costi e delle risorse impiegate in ambito di *speech technology*: qualora tali effetti non fossero universali, ma specifici per lingua o cultura, le società con simili interessi sarebbero obbligate a personalizzare gli algoritmi. A tal riguardo l'assunto di base di Scherer sta nel fatto che le inferenze dei giudici ascoltatori siano basate proprio sui parametri acustici che determinano il profilo delle varie emozioni. Sebbene in studi precedenti (Banse & Scherer, 1996) Scherer affermi che gli effetti delle emozioni sulla voce debbano essere universali, lo stesso autore aggiunge:

*Unfortunately, there is no study to date that examined these effects across speakers from different languages and culture.*¹³⁷

Inoltre, nonostante i dati supportino l'assunto teorico che il riconoscimento sia in gran parte guidato da meccanismi psicobiologici universali (dal momento che soggetti di cultura e

¹³⁷ Cfr. Scherer (2000), pag. 380.

lingua diversa riconoscono le emozioni espresse con un'accuratezza superiore al caso), aggiunge anche come non appena vengano utilizzati elementi diversi da *pure nonlinguistic affect bursts*, fattori legati ad aspetti segmentali e soprasegmentali delle diverse lingue influenzano sia l'*encoding* che il *decoding* delle emozioni, concludendo come gli effetti debbano essere:

“[...] either due to segmental information (such as phoneme specific fundamental frequency, articulation differences, formant structure, or the like) or to suprasegmental parameters, such as prosodic cues (intonation, rhythm, timing).”¹³⁸

Per le difficoltà e gli elementi emersi nella sua indagine, che lasciano inevitabilmente aperti e senza risposta molti dei dubbi iniziali, Scherer (2000: 382) conclude affermando:

“further research using encoders and decoders from different languages and cultures is urgently needed to establish in greater detail which language or culturespecific expression rules speech technology needs to incorporate into the development of products to be marketed on a global scale.”

Oltre allo studio appena citato se ne potrebbero naturalmente riportare molti altri che, puntando sull'elaborazione delle caratteristiche emotive del parlato, trovano grande applicazione nel riconoscimento automatico del parlato (*automatic speech recognition*), nella sintesi del parlato da testo scritto (*text-to-speech synthesis*) o ancora nelle oramai diffusissime *facce parlanti*.

5.1 *Facce parlanti: punto di contatto tra emozioni facciali e vocali*

È stato più volte ribadito come le emozioni facciali e le emozioni vocali rappresentino due ambiti assai differenti, sia per tipologia di approccio, sia per tipologia di segnale. A fronte della diversità di questi due tipi di segnale, quello facciale e quello vocale, vi sono, allo stesso tempo, anche una serie di punti di contatto, di elementi comuni a cui si è più volte accennato, che nel contesto qui descritto trovano larghissima applicazione in quelle che oggi vengono comunemente chiamate *facce parlanti*.

Che quello delle *facce parlanti* sia un ambito in costante evoluzione che ha avuto, a livello applicativo e commerciale, un grande successo è indubbio (si pensi, ad esempio, alle numerose produzioni cinematografiche e di animazione, dove in termini di dettaglio e realistica si sono raggiunti livelli inimmaginabili fino a qualche anno fa).

¹³⁸ *Ibidem*, pag. 382.

Affrontare qui una tematica del genere esula certamente da quelli che sono gli scopi della ricerca intrapresa. Ciò su cui si vuole però porre l'accento riguarda invece l'altro settore delle facce parlanti che nasce principalmente come ambito di applicazione di risultati di precedenti ricerche che mirano per lo più a fornire una corretta caratterizzazione, sia in termini articolatori che percettivi, di elementi utili alla generazione di *facce parlanti*. Le finalità, oltre ad essere connesse al miglioramento dei tradizionali sistemi di riconoscimento basati sull'informazione ottenibile attraverso il solo canale acustico, sono orientate a rendere l'interazione uomo-macchina più *human-like* e quanto più naturale possibile. Quelle appena menzionate sono tematiche sviluppate da studiosi provenienti da ambiti scientifici diversi che hanno sicuramente avuto dalla loro parte il crescente miglioramento delle strumentazioni che risultano oramai contraddistinte da elevatissime capacità computazionali.

Quanto fin ora descritto fa già parte del presente. Basta guardarsi intorno o fare una semplice ricerca sul web per rendersi conto che, a parte quelle che abbiamo definito del primo tipo e che esulano dagli obiettivi qui perseguiti, esistono *facce parlanti* di ogni natura con caratteristiche che si possono definire di:

- a. ecologicità, in quanto riproducono condizioni di interazione faccia-a-faccia trasmettendo informazione uditiva e visiva;
- b. naturalezza, ovvero riproducono la fusione di informazioni linguistiche, paralinguistiche ed extralinguistiche;
- c. robustezza, la comunicazione bimodale (uditiva e visiva) assicura la comprensibilità del messaggio in condizioni ambientali che degradano il segnale acustico con ricadute positive su soggetti con deficit di tipo uditivo.¹³⁹

Sono queste le tre caratteristiche che offrono alle *facce parlanti* grandi possibilità applicative. Si pensi, ad esempio, alla faccia parlante *BALDI*, un *language tutor* attualmente commercializzato, interamente integrato in uno *speech toolkit* in grado di fornire concreto aiuto nell'apprendimento del linguaggio a bambini affetti da sordità, a bambini autistici e a bambini con ritardi nello sviluppo del linguaggio. Dietro a tutto ciò si nascondono una serie di studi di tipo articolatorio e percettivo.¹⁴⁰

¹³⁹ Cfr. Magno Caldognetto, E., Cavicchio, F., Cosi, P., Drioli, C., Tisato, G. (2005), Parametri per lo studio delle modificazioni articolatorie nel parlato emotivo, in Cosi P. (a cura di), *CD-Rom Proceedings of AISV 2004, 1st Conference of Associazione Italiana di Scienze della Voce, Padova, Italy, December 2-4, 2004*, Padova, EDK Editore s.r.l., pag. 449.

¹⁴⁰ Cfr. Massaro, D. W., Cole, R. (2000), From Speech is special to talking heads in language learning, in: *Proceedings of Integrating Speech Technology in the (Language) Learning and Assistive Interface, (InSTIL 2000)*, University of Abertay Dundee, Scotland, pp. 153-161.

Vi è poi un filone di ricerca interamente italiano che si è occupato dello studio della correlazione tra informazione trasmessa dal messaggio verbale, caratteristiche acustiche della voce emotiva ed espressione facciale co-prodotta alle emozioni, con tecniche optoelettroniche per la raccolta di dati cinematici allineati appunto a dati articolatori ed acustici.¹⁴¹

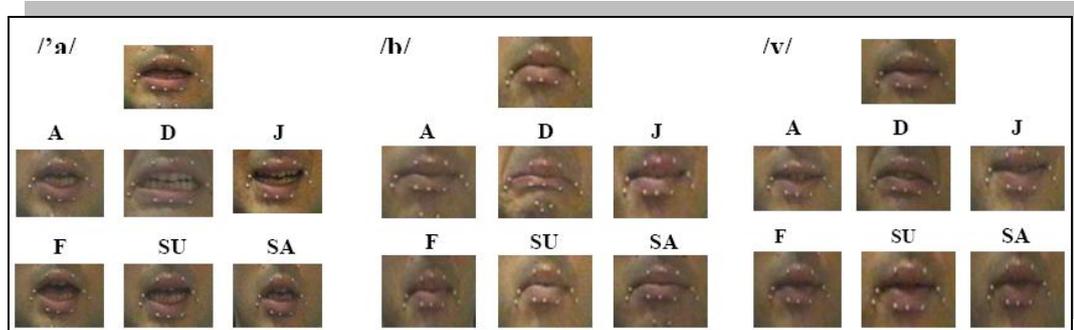


Figura 28: Un esempio dello studio delle configurazioni labiali emotive relative al punto medio dei bersagli articolatori /a, b, v/ per le emozioni di anger, disgust, joy, fear, surprise e sadness¹⁴²

Ciò che in questo caso risulta particolarmente rilevante viene dal fatto che dall'insieme degli studi condotti si è potuti giungere alla definizione e creazione di un software denominato *INTERFACE*¹⁴³ che consente di interagire sia con *Lucia*¹⁴⁴ per la creazione di espressioni facciali emotive, sia di generare un parlato emotivo in cui i movimenti articolatori e labiali vengono coprodotti a quelli delle espressioni facciali.¹⁴⁵

5.1.1 La codifica delle emozioni facciali attraverso il FACS

Avendo introdotto l'argomento delle *facce parlanti* non si può non affrontare il discorso del FACS (*Facial Action Coding System*), un sistema per l'*encoding* o la codifica delle espressioni facciali.¹⁴⁶ Si tratta di un sistema di codifica a base anatomica, che divide il viso in unità muscolari dette AUs (*Action Units*) che per la sua capacità di codificare tutti i possibili movimenti del viso rappresenta attualmente il sistema di codifica di riferimento. Le AUs descrivono, infatti, l'attività muscolare sottostante alle diverse possibili modificazioni espressive della faccia che possono essere apprese e riconosciute.

¹⁴¹ Limitandoci alla lingua italiana si vedano ad esempio gli studi condotti da Magno Caldognetto *et al.* (2003a), Magno Caldognetto *et al.* (2003b), Magno Caldognetto *et al.* (2004a), Magno Caldognetto *et al.* (2004b), Magno Caldognetto *et al.* (2005).

¹⁴² Figura tratta ed adattata da Magno Caldognetto *et al.* (2005), pag. 451.

¹⁴³ Cfr. Tisato, G., Cosi, P., Drioli, C., Tesser, F. (2005), *INTERFACE: a New Tool for Building Emotive/Expressive Talking Heads*, in Proc. INTERSPEECH 2005, Lisbon, Portugal, pp. 781-784.

¹⁴⁴ Cfr. Cosi, P., Fusaro, A., Tisato, G. (2003), *LUCIA a New Italian Talking-Head Based on a Modified Cohen-Massaro's Labial Coarticulation Model*, Proc. Eurospeech 2003, Geneva, Switzerland, pp. 127-132.

¹⁴⁵ Per una analisi più completa e dettagliata si rimanda alle pagine web dedicate, rispettivamente <http://www2.pd.istc.cnr.it/INTERFACE/home/> per *INTERFACE*, <http://www2.pd.istc.cnr.it/Lucia/home/> per *Lucia*.

¹⁴⁶ Il FACS nasce dagli studi di Ekman sul *decoding* delle espressioni facciali in diverse culture.

Il motivo per il quale si accenna qui a questo sistema di codifica delle espressioni facciali deriva dalla possibilità di descrivere anche la regione delle labbra che, come più sopra accennato, rappresenta uno degli articolatori mobili fondamentali per la produzione dei suoni della parola, senonchè l'unico degli articolatori visibili in quanto posto nella parte esterna e visibile di quello che costituisce il nostro volto.

Riguardo la sola regione delle labbra, le azioni più importanti che producono trasformazioni nella forma delle labbra sono quelle che: le allungano (*elongate*), le accorciano (*de-elongate*), le assottigliano (*narrow*, con riferimento alla parte rossa delle labbra), le ingrandiscono (allargano, *widen*: sempre in riferimento alla parte rossa delle labbra), le appiattiscono (*flatten*), le protrudono (*protrude*), le contraggono (*tighten*), le stirano (*stretch*), le arricciano in dentro (*roll in*). Ciascuna di queste azioni è anche correlata ad una scala di visibilità che corrisponde a punteggi progressivi di intensità (*slight, marked, pronounced, severe, extreme, maximum*) in cui gli intervalli non sono eguali e le differenze sono comunque molto ridotte, dato che l'insieme dei movimenti coinvolti è limitato spazialmente.¹⁴⁷

L'EMFACS (*Emotion FACS*) è un sottoinsieme del sistema di codifica delle espressioni facciali che prevede l'utilizzo delle sole AUs o dei muscoli utili alla rappresentazione delle emozioni di cui viene di seguito fornita una esemplificazione grafica.

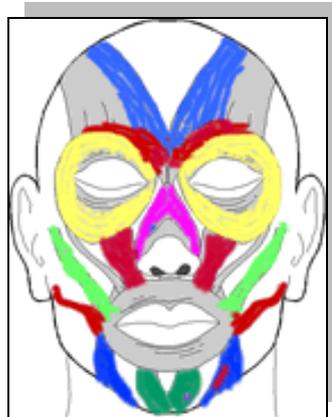


Figura 29: AU o fasce muscolari che compongono le EMFACS¹⁴⁸

5.1.2 Affective Presentation Markup Language (APML)

L'*Affective Presentation Markup Language* (APML) è un linguaggio di *mark-up* emotivo che, applicato ai testi, permette di definire le regole che determinano le caratteristiche della sintesi vocale e bimodale, nonché uditivo-visiva, in una *faccia parlante* o in agenti

¹⁴⁷ Per una descrizione più completa con le AUs specificamente interessate si veda Magno Caldognetto *et al.* (2005), pp. 456-457.

¹⁴⁸ Immagine tratta da http://www.face-and-emotion.com/dataface/facs/media/facmusc_col.gif (data consultazione: 22 maggio 2008).

conversazionali virtuali. Si tratta di un linguaggio di annotazione sostanzialmente basato e derivato dal linguaggio XML (*eXtensible Markup Language*) che, date le sue caratteristiche flessibili, rappresenta allo stato attuale uno dei modi più diffusi per aggiungere informazioni ad un testo scritto arricchendolo.¹⁴⁹

Attraverso il linguaggio APML è quindi possibile definire e codificare nel testo, attraverso l'utilizzo di specifici *tags*¹⁵⁰, una serie di funzioni comunicative da sintetizzare e riprodurre in una *faccia parlante*, funzioni comunicative che in Poggi (2003) vengono definite come associazione di significato-segnale. In Bevacqua *et al.* (2008), ad esempio, tali funzioni comunicative¹⁵¹ vengono utilizzate per controllare il comportamento di un agente virtuale in ambito di *e-learning* per migliorarne le abilità socio-emotive.

In modo del tutto analogo, all'interno del modulo denominato *Emotion Player* di *INTERFACE*, è stata implementata un'interfaccia che, da un lato, offre la possibilità di generare un file di testo in linguaggio APML, dall'altro, di rappresentare le funzioni comunicative riscontrabili in una comune conversazione.¹⁵²

5.2 Emozioni, Intelligenza Artificiale ed Affective Computing

Sebbene si tratti di aspetti che meritano di essere approfonditi nella giusta maniera, anche in questo caso ci si limiterà solo a qualche accenno per evitare di essere sviati dagli obiettivi primari della presente ricerca.

Come sottolinea Gola (2002) nel suo capitolo "L'approccio alle emozioni in IA", anche in ambito di Intelligenza Artificiale (IA) le emozioni iniziano a riscuotere un enorme interesse il cui scopo primario è quello di dotare i diversi sistemi computazionali di funzioni e abilità emotive.

¹⁴⁹ Per maggiori dettagli *cfr.* De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M. (2003), APML, a mark-up language for believable behavior generation, in: Prendiger, H., Ishizuka, M. (Eds.), *Life-like characters. Tools, affective functions and applications*, Berlin: Springer-Verlag, pp. 65-86.

¹⁵⁰ In un qualsiasi linguaggio di annotazione il *tag* è un elemento che definisce una specifica caratteristica di una porzione di testo e viene posto rispettivamente prima e dopo la porzione di testo interessata.

¹⁵¹ Le funzioni comunicative proposte da Poggi (2003) vengono riassunte in Bevacqua *et al.* (2008: 132) nel seguente modo:

1. informazione sulle convinzioni dello speaker: comportamenti che danno informazioni sulle convinzioni di chi parla, ad esempio sul grado di certezza di quello che si sta dicendo;
2. informazione sulle intenzioni dello speaker: chi parla può dare informazioni sui suoi scopi attraverso, ad esempio, la scelta dei gesti o mettendo in risalto una certa parte della frase;
3. informazione sullo stato emotivo dello speaker: chi parla comunica il proprio stato emotivo attraverso le espressioni facciali;
4. informazione metacognitiva: chi parla può comunicare il fatto che sta cercando di ricordare un'informazione.

¹⁵² *Cfr.* Magno Caldognetto *et al.* (2008: 209).

Nell'immaginario comune l'interazione tra uomo e calcolatore viene generalmente immaginata come un processo in cui il primo impartisce comandi, mentre il secondo li esegue. Non mancano, tuttavia, i casi in cui i ruoli si invertono, al punto che è l'uomo ad eseguire le istruzioni impartite dalla macchina.¹⁵³ Sebbene l'uomo sia in grado di eseguire una serie di disposizioni sottoforma di algoritmi, non sempre è sufficiente presentargli delle istruzioni con l'obiettivo che lui le esegua correttamente. Affinché ciò sia possibile, è necessario che l'uomo comprenda, prima di tutto, il contenuto di ogni istruzione che gli viene impartita e che ad essa corrisponda effettivamente la capacità da parte sua di eseguirle. Come è facile aspettarsi anche lo stato emotivo e cognitivo di una persona possono, infine, influire significativamente sulle capacità che egli ha di eseguire determinate istruzioni. Si tratta di alcune considerazioni che sino a poco tempo fa non erano tanto tenute in conto: le conseguenze di ciò si riflettevano, infatti, in procedure alquanto rigide e poco adattate all'utente o alla persona a cui venivano impartite le istruzioni.

Partendo da tali considerazioni e grazie soprattutto ai progressi della ricerca scientifica sulle emozioni è stata intravista la possibilità di trattare simili situazioni in ambito computazionale facendo nascere e dando origine ad un nuovo ambito di ricerca a cui è stato dato il nome di *affective computing*. L'*affective computing* è perciò quel campo di investigazione che si occupa del trattamento computazionale delle emozioni che ha alla base lo scopo di dotare un computer di abilità affettive, con la

*“capacità di riconoscere, esprimere e ‘avere’ emozioni, in un senso simile a quanto accade negli esseri umani”.*¹⁵⁴

Fino a non molti anni fa una affermazione del genere si sarebbe rivelata del tutto improponibile in quanto la contemplazione della sfera emotiva nelle macchine veniva considerata elemento di disturbo per i processi intelligenti con effetti catastrofici che li avrebbero resi inefficienti, imprevedibili e incontrollabili dall'essere umano.¹⁵⁵

In tale ambito Rosalind Picard, che dirige il gruppo di ricerca sull'*Affective Computing* presso il Massachusetts Institute of Technology (MIT), è una delle maggiori esponenti di questa recente disciplina. Il suo scritto *Affective Computing* (Picard, 1997) viene considerato

¹⁵³ È il caso, ad esempio, di alcune funzionalità di tipo *task-oriented* fornite da sistemi di aiuto in linea con la presentazione all'utente di sequenze di istruzioni necessarie ad eseguire un'azione complessa.

¹⁵⁴ Cfr. Gola, E. (2002), pag. 185.

¹⁵⁵ Uno stereotipo assai diffuso di questo stato delle cose si può riscontrare nel caso di HAL nel film del 1968 “2001: Odissea nello spazio” (regia di Stanley Kubrick).

il testo di riferimento più importante e ufficiale su quello che è il tema e le relazioni tra emozioni ed IA e al quale ci si affida per questioni connesse alla modellizzazione emotiva.¹⁵⁶

¹⁵⁶ *Cfr.* Gola, E. (2002), pag. 185, a cui si rimanda per tutto ciò che in questa sede avrebbe meritato di essere trattato e che è stato evitato per le motivazioni a cui si è più volte accennato.

6 Questioni metodologiche nello studio delle emozioni vocali

Alle ricerche sul parlato emotivo si aggiungono, oltre ad una serie di problematiche simili a quelle delle espressioni facciali, tutta una serie di altre difficoltà, difficoltà che, nel caso di ricerche a livello cross-linguistico, vengono ulteriormente complicate in termini metodologici.

Una delle più grandi e significative differenze tra la ricerca sulle emozioni trasmesse attraverso le espressioni facciali e quella sulle emozioni comunicate tramite espressioni verbali è data soprattutto dalla diversità di approccio che lo studio in questione richiede, differenza che Pell *et al.* (2009: 108) sintetizzano nel seguente modo:

“Contrary to research on the face, researchers interested in the voice cannot present a valid ‘snapshot’ which represents the vocal attributes of an emotion.”¹⁵⁷

Si tratta di una differenza per nulla banale. Come fanno notare Poggi & Magno-Caldognetto (2004), il compito di chi si è occupato di parlato emotivo è risultato e risulta ancor più complicato a causa della grande variabilità dei dati raccolti, siano essi di tipo intra-linguistico che di tipo inter-linguistico, variabilità dovuta alle limitazioni metodologiche (tipo di parlato spontaneo *vs.* parlato di laboratorio, letto o recitato); alla scelta dei parlanti (attori *vs.* parlanti naif); al metodo di elicitazione delle emozioni (fotografie, scenari, etichette linguistiche ecc.) e, di non minore importanza, alla scelta del materiale linguistico prodotto (vocali, sillabe, interiezioni, stringhe fonologiche non-sense ecc.). Risulta quindi opportuno, nei paragrafi che seguiranno, mettere in evidenza alcune questioni di ordine metodologico e relative problematiche connesse, sulla base delle quali è stato successivamente possibile adottare determinate strategie nello sviluppo del progetto e della ricerca qui proposta.

¹⁵⁷ Cfr. Pell, M. D., Monetta, L., Paulmann, S., Kotz, S. A. (2009), Recognizing emotions in a foreign language, in: *Journal of Nonverbal Behavior*, 33, pag. 108.

6.1 Limitazioni di tipo etico e morale

Contrariamente all'opinione comune, anche nella ricerca scientifica sulle emozioni esistono limiti di tipo etico e morale. In tal senso sono, infatti, numerosi i ricercatori che sino ad oggi si sono adoperati per riuscire a superare i limiti imposti da questioni di natura etica, qualunque fosse la motivazione dietro allo studio delle emozioni, come sottolineano a questo proposito Rottenberg *et al.* (2007: 9), quando scrivono:

“whatever their motivation for studying emotion, experimentalists have required a reliable means of eliciting emotion in an ethically acceptable fashion.”

La questione etica non è, quindi, cosa di poco conto. Nel mondo occidentale, ma oramai ovunque, la ricerca scientifica prevede che la stessa si conformi e rispetti specifiche linee guida di tipo etico e morale. Il ricercatore non può ricorrere a tutti gli stratagemmi che gli vengono in mente e non può in ogni caso indurre nei soggetti sperimentali emozioni o sensazioni oltre quelle che ci si aspetterebbe nella vita quotidiana da parte di soggetti non affetti da disturbi o psicopatologie a carico della sfera affettiva. Allo stesso modo, le emozioni indotte sperimentalmente dovrebbero poter essere indotte facendo ricorso a stimoli o situazioni comunemente riscontrabili nella vita di tutti i giorni. Non a caso, l'induzione di emozioni fortemente dolorose viene spesso limitata o addirittura rifiutata. Anche quando il ricorso all'induzione di emozioni negative dovesse rendersi assolutamente necessario, le stesse emozioni indotte dovrebbero essere transitorie, ovvero devono estinguersi non appena il soggetto abbia abbandonato il laboratorio e in ogni caso non devono lasciare segni nel soggetto utilizzato nella sperimentazione.

Come rilevano Niedenthal *et al.* (2006), le limitazioni etiche e morali non sono sempre state così evidenti, al punto che negli anni '50 del secolo scorso fu condotto un esperimento che oggi sarebbe assolutamente improponibile. Ax (1953), in una ricerca il cui scopo era quello di dimostrare che specifici *patterns* dell'attività del sistema nervoso autonomo fossero in grado di caratterizzare emozioni discrete come gioia, paura e rabbia, mise in atto una sorta di inganno ai danni dei soggetti che presero parte all'esperimento. Con la scusa di testare la validità del poligrafo, Ax collegò gli ignari soggetti ad una serie di dispositivi, tra cui un generatore di elettroshock, al contempo registrando e misurando ben 14 indicatori dell'attività

del Sistema Nervoso Autonomo¹⁵⁸. Durante l'esperimento, per l'emozione della paura, il ricercatore simulò una serie di malfunzionamenti provocando delle scintille all'apparecchio dell'elettroshock facendo insorgere, nei soggetti a cui lo stesso apparecchio era collegato, un alto livello di angoscia e momenti di intensa paura. Per la situazione di collera, invece, il ricercatore costrinse i partecipanti a ben 5 minuti di abusi da parte di un operatore maldestro suscitando nei soggetti livelli di fastidio e di collera molto intensa, dimostrando, per l'appunto, quelle che erano le sue ipotesi.¹⁵⁹

6.2 “Catturare” le emozioni

Disporre di procedure di elicitazione valide e funzionali agli scopi della ricerca che si intende intraprendere, e da cui ricavare il materiale oggetto di studio e di analisi, rappresenta, anche in virtù delle limitazioni sopra esposte, un obiettivo di primaria importanza a cui molti ricercatori e studiosi interessati alle emozioni nei vari ambiti, si sono impegnati nel tentativo di mettere a punto procedure che fossero quanto più possibile controllate e replicabili.

Sebbene il recente *Handbook of Emotion Elicitation and Assessment*, curato da James A. Coan e John J. B. Allen vanta una ricca ed esaustiva raccolta delle metodologie più comuni per l'elicitazione e l'induzione di stati affettivi/emotivi in laboratorio¹⁶⁰, la raccolta o elicitazione delle emozioni non è per niente cosa facile per il fatto che, come affermano gli stessi curatori del volume,

“no single sensory domain is paramount – that indeed emotions can arise through a variety of modalities and that these modalities (even some that can be controversial) need to be well understood and properly implemented [...]”¹⁶¹

Oltre ad evidenziare le difficoltà connesse a questo aspetto, appare evidente come si tratti, spesso ed esclusivamente, di tecniche sperimentali sulle quali non esiste allo stato attuale grande consenso. Da più parti si assiste, infatti, ad un'impressionante proliferazione di nuove tecniche di elicitazione che, spesso e volentieri, non trovano sufficiente spazio per i dovuti approfondimenti, e che per questo motivo vengono relegate a sommarie descrizioni:

“Unfortunately, the published literature offers little explicit on these issues, forcing investigators to base their decisions on

¹⁵⁸ Tra cui pressione del sangue, battito cardiaco, respirazione, tensione muscolare ecc..

¹⁵⁹ Cfr. Niedenthal *et al.* (2006), pag. 23.

¹⁶⁰ Il volume dedica, infatti, sotto il nome di “Emotion elicitation”, una intera sezione di 10 capitoli a quelle che sono le tecniche di elicitazione più diffuse e conosciute.

¹⁶¹ Cfr. Coan, Allen (2007), pag. 4.

*informal rules of thumb, idiosyncratic training experiences, or personal communications.*¹⁶²

Ciò risulta ancor più vero se si considera quanto già richiamato nell'introduzione a questo lavoro che evidenzia come, da un lato, quello delle scienze affettive rappresenta oggi un dominio di ricerca non più associato alla sola psicologia, ma investe e chiama in causa una serie di altre discipline, all'interno delle quali, ciascuno dei suoi appartenenti tenta di dare il proprio contributo attraverso le proprie tecniche e le proprie conoscenze; dall'altro, ed è questo l'aspetto che appassiona e intriga di più, le emozioni, intese nell'accezione più generica del termine, investono e influenzano ogni livello di indagine chiamandone in causa altri. Lo studio delle emozioni come tale, oltre a richiedere specifiche competenze che abbracciano più di una disciplina, richiede disegni sperimentali e strategie ben definite, tecniche di indagine e strumentazioni anche molto sofisticate dove la collaborazione tra ricercatori e laboratori diventa essenziale, se non addirittura inevitabile. Lo studio delle emozioni richiede, infatti,

*“attention paid to multiple levels of analysis, from the cultural to the behavioral, psychological, experiential, physiological, and molecular”.*¹⁶³

Nell'ambito dello studio delle emozioni espresse attraverso la voce, le strategie e le tecniche di raccolta del parlato emotivo si differenziano tra loro per tipo di approccio utilizzato e numero di emozioni che sono in grado di indurre.

La scelta di una metodologia, piuttosto che l'altra, è strettamente connessa alle finalità e al disegno sperimentale della ricerca che si intende intraprendere, con la consapevolezza che la stessa sarà caratterizzata da vantaggi e svantaggi che vanno tenuti in debita considerazione. Dal momento che non esistono procedure standardizzate per la raccolta di parlato emotivamente connotato, anche il tentativo di una suddivisione di quelle che possono essere le tecniche utilizzate diventa assai difficile, se non addirittura fuorviante, proprio perché molte di queste si compenetrano e si completano a vicenda. Quanto appena detto risulta chiaramente sintetizzato in un breve passaggio che di seguito viene riportato:

¹⁶² Cfr. Rottenberg *et al.* (2007), pag. 10. Poco più avanti gli autori riconducono le difficoltà connesse alle procedure di elicitazione e/o induzione di emozioni al fatto che la definizione stessa di emozione sia, di per sé, storicamente assai controversa: “*With different investigators employing their own idiosyncratic and often widely discrepant definitions of emotion, it is no wonder that there has been considerable confusion as to which procedures reliably elicit emotion.*” Ciò spiegherebbe, di fatto, l'assenza di procedure standardizzate per la raccolta di materiale emotivo (voce o altro che sia).

¹⁶³ *Idem*, pag. 6.

“Because emotion is a multifaceted about which so much is still unknown, no single technique can serve all purposes for eliciting emotion in the laboratory. Thus the very same properties of a given induction procedure that make it a valuable procedure to probe emotion in many contexts may represent liabilities in others.”¹⁶⁴

Un altro problema a cui alcuni studiosi si sono dedicati, riguarda la possibilità di stabilire quanto effettivamente l'utilizzo di una procedura di elicitazione sia efficace. Vi sono innumerevoli studi che evidenziano differenze sostanziali tra le varie metodologie in riferimento alla loro efficacia: si va, infatti, da un 87% di riuscita per tecniche che fanno uso della musica come strumento di elicitazione, per piombare ad un bassissimo 15% per tecniche che fanno uso di suggestione ipnotica.¹⁶⁵ In una meta-analisi condotta su 138 studi da Westermann *et al.* (1996), invece, la presentazione di film o storie si è rivelata essere la tecnica più efficace nell'induzione stati affettivi sia positivi che negativi.

Sebbene una chiara suddivisione di quelle che sono le tecniche utilizzate sia impossibile, in letteratura sono presenti procedure che producono:

- parlato emotivo “indotto” sperimentalmente con l'ausilio di *Mood Induction Procedures* (MIPs) che utilizzano particolari procedure di laboratorio a cui il soggetto viene sottoposto, e che producono materiale emotivo che si potrebbe definire “semi-naturale”;
- parlato emotivo “naturale”, e quindi spontaneo e autentico, ricavato da situazioni di vita quotidiana o frutto di particolari interazioni uomo-uomo o anche uomo-macchina;¹⁶⁶
- parlato emotivo “simulato”, o recitato, anche qui secondo schemi e modalità predeterminate.¹⁶⁷

Tenendo presenti i limiti relativi alla raccolta di materiale emotivo descritti in questo paragrafo, nel prosieguo si tenterà di fornire una visione il più possibile esauriente di quelle che sono le tecniche più comunemente utilizzate per la raccolta di produzioni emotive,

¹⁶⁴ Cfr. Rottenberg *et al.* (2007), pag. 11.

¹⁶⁵ Cfr. Eich *et al.* (2007), pag. 125.

¹⁶⁶ Sulla possibilità di utilizzo di parlato di questo tipo sussistono questioni legate alla privacy dei soggetti registrati e che rientrano in quelle che sono state definite nel precedente § come “*Limitazioni di tipo etico e morale*”. Nonostante ciò si assiste, allo stato attuale, ad un crescente ricorso a parlato emotivo “reale” proprio per le opportunità e per le possibilità che questo tipo di materiale offre in applicazioni commerciali *real-time*, come quella in forte crescita dei servizi telefonici.

¹⁶⁷ La presente tassonomia, relativa ai tre tipi di parlato generalmente raccolto e riscontrabile negli studi sulle emozioni vocali, fa riferimento a quella riportata da Scherer (2003) a pag. 231 e segg. Per quel che riguarda la raccolta di produzioni emotive relativamente alle tre tipologie a cui qui si fa riferimento, diversi ricercatori hanno, come è facile intuire, perseguito l'obiettivo in modo diverso.

sottolineando il fatto che molte di esse vengono fatte interagire rendendo in alcuni casi assai difficile la tripartizione a cui più sopra si accennava.¹⁶⁸

Nella tripartizione che seguirà sarà inoltre utile ricordare come la tipologia di parlato (letto vs. spontaneo ecc.), implicito nelle procedure di raccolta delle emozioni a cui si accennerà, rappresenta una ulteriore limitazione e/o possibilità di raccogliere dati aventi determinate caratteristiche e che Fagyal (1995) rappresenta secondo due scale di valutazione: una relativa livello di autenticità delle produzioni (che va da non spontaneo, come la lettura, a spontaneo, come i resoconti verbali di esperienze vissute), e una relativa al livello di controllo del *setting* sperimentale (che va dalla possibilità di controllo assoluto alla totale impossibilità di controllo nelle procedure di raccolta).

AUTHENTICITE		CONTROLE EXPERIMENTAL	
Non-spontané	0	<u>LECTURE</u>	↑ contrôlé
	1	Logatomes	12
	2	Mots	11
	3	Phrases 'réitérées'	10
	4	Phrases 'probabilitaires'	9
	5	Phrases ordinaires	8
	6	Textes	7
Semi-spontané		<u>DIALOGUE</u>	Semi-contrôlé
	7	Question-Réponses prédéfinies	6
	8	Questions avec réponses non définies	5
	9	Description d'image	4
	10	Interprétation d'image	3
	11	Conversation dirigée	2
	12	Rappel de mémoire	1
Spontané		0 Non-contrôlé

Figura 30: Scala di autenticità e di controllo sperimentale sull'acquisizione di produzioni vocali secondo Fagyal (1995: 46)

Come si evince dalla Figura 30 qui riportata, il livello di autenticità è inversamente proporzionale al livello di controllo del materiale raccolto.

Tenendo a mente la schematizzazione appena fornita, verrà di seguito affrontata la questione delle possibili modalità di raccolta di parlato emotivo secondo la tripartizione più sopra fornita.

¹⁶⁸ Va altrettanto evidenziato come alcune delle tecniche che di seguito verranno presentate siano tecniche utilizzate anche nello studio delle espressioni facciali.

6.2.1 Emozioni indotte: le *Mood Induction Procedures* (MIPs)

Una delle modalità di raccolta di materiale emotivamente connotato è data dall'utilizzo di tecniche genericamente identificate con il termine di *Mood Induction Procedures* (MIPs). Le MIPs sono per lo più tecniche o strategie di laboratorio,

*"whose aim is to provoke in an individual a transitory emotional state in a non natural situation and in a controlled manner; the mood induced tries to be specific and pretends to be an experimental analogue of the mood that would happen in a certain natural situation."*¹⁶⁹

Con riferimento alle MIPs, Eich *et al.* (2007: 125) sottolineano come tecniche sperimentali di questo tipo siano state messe a punto per prime e come siano in forte aumento e diventino sempre più popolari. Data quest'ultima peculiarità, e ipotizzando che questo trend continui a crescere, si chiedono:

*"What can be done to maximize the benefits of experimental mood modification, while minimizing its costs?"*¹⁷⁰

La risposta si riassume in quella che loro stessi definiscono una *wishlist* di sei punti a cui una MIP dovrebbe rispondere e che qui di seguito viene riprodotta in Tabella 8:

Desirable attributes of a mood induction technique

- Technique has a high rate of success in altering participants' moods in predictable ways.
 - Technique allows for individual differences in time taken to develop a particular mood.
 - Induced moods are strong or intense.
 - Induced moods are stable over time and across tasks.
 - Induced moods seem real or authentic to the participants.
 - One and the same mood can be reliably induced on more than one occasion.
-

Tabella 8: *Wishlist* degli attributi che una tecnica di *mood induction* dovrebbe possedere¹⁷¹

Esempi di induzione di emozioni prevedono tecniche che fanno uso di sequenze video o di film di cui viene fornita una accurata descrizione in Rottenberg *et al.* (2007). Altri propongono come MIPs particolari tecniche di realtà virtuale, come ad esempio Baños *et al.* (2006) che utilizzano un parco virtuale che cambia in relazione all'umore da indurre. Altri ancora hanno fatto uso di brani musicali, tecnica utilizzata per la prima volta da Sutherland, Newman, Rachman (1982), che hanno chiesto a dei soggetti di utilizzare i brani musicali presentati come mezzo per entrare in uno stato d'animo, avvertendoli allo stesso tempo del

¹⁶⁹ Cfr. García-Palacios, Baños (1999), pag. 16, *op. cit.* in Baños *et al.* (2006).

¹⁷⁰ Si tratta in realtà di una domanda che ben si presta a qualsiasi tecnica di elicitazione e/o induzione di stati affettivi.

¹⁷¹ Tabella proposta (ed adattata) da Eich *et al.* (2007), pag. 125, tab. 8.1.

fatto che la musica da sola non è in grado di indurre in loro lo stato d'animo in modo automatico, ma che dovevano fare ricorso a proprie strategie per raggiungerlo. Alcuni si sono avvalsi di specifiche tecniche di induzione per valutarne gli effetti sulla manipolazione della voce, come ad esempio Johnstone *et al.* (2005) che hanno utilizzato un videogame per l'induzione di espressioni emotive naturali in laboratorio. Altri ancora hanno studiato il parlato emotivo spontaneo in conversazioni di vita reale, o in situazioni di elevato stress emotivo, come ad esempio in una ricerca condotta da Laukka *et al.* (2008) in cui sono state studiate espressioni vocali autentiche¹⁷² raccolte con due diverse tecniche: a) registrazione di parlato di soggetti affetti da fobia sociale durante un colloquio pubblico ansiogeno prima e dopo trattamento; b) registrazione di parlato frutto di interazioni uomo-macchina in condizioni reali. Ulteriori tecniche per indurre determinate emozioni, oggi sicuramente meno attuabili per le ragioni a cui più sopra si è accennato, prevedono l'utilizzo di droghe come in Helfrich *et al.* (1984) in cui sono stati studiati gli effetti di antidepressivi sui diversi parametri vocali per un periodo di diverse ore.

I vantaggi derivanti dall'utilizzo delle MIPs sono dati dal fatto che si riescono ad ottenere emozioni pressoché spontanee con un controllo assoluto della situazione di laboratorio appositamente creata; al contrario, nella maggior parte dei casi le emozioni prodotte risultano però difficili da identificare o da etichettare per l'elevata variabilità che caratterizza le produzioni da un soggetto all'altro. Tuttavia, come già accennato, accanto alla possibilità che queste tecniche offrono nell'indurre emozioni "reali", a seconda dei casi sussistono limitazioni di tipo etico che difficilmente sono superabili.

Pro	Contro
Emozioni spontanee e "reali"	Difficoltà di identificazione e/o etichettatura
Elevata ecologicità	Elevata variabilità
Controllo sulle procedure di induzione	Limitazioni di tipo etico

Tabella 9: Tabella riassuntiva degli elementi a favore o a sfavore delle MIPs

A questa tipologia appartiene anche il metodo cosiddetto "Velten" (Velten, 1968), una tecnica di MIP per l'induzione di stati affettivi positivi e/o negativi attraverso il comportamento verbale. Nella sua versione originaria¹⁷³ la procedura prevede che il metodo venga somministrato oralmente e individualmente ai partecipanti a cui viene chiesto di leggere una lunga lista di affermazioni emotivamente caratterizzate, dapprima a se stessi e poi

¹⁷² Basandosi principalmente su espressioni *non-basic*, ovvero *anxiety, irritation, resignation*.

¹⁷³ Ne esiste anche una versione rivista adottata da Seibert, Ellis (1991) in cui, rispetto alla versione di Velten (1968), gli *item* della procedura contengono espressioni di uso familiare e di uso quotidiano tipico degli studenti, non contengono riferimenti a stati d'animo di tipo somatico e l'intero set si presenta con solo 25 produzioni invece di 60.

ad alta voce. Le affermazioni somministrate, che vanno da uno stato inizialmente neutro per arrivare a connotazioni emotive positive o negative a seconda dei casi presentati¹⁷⁴, sono così in grado di produrre cambiamenti nell'umore del soggetto che le legge.

In un esperimento condotto da Wilting, Kraemer, Swerts (2006) il metodo “Velten” è stato, ad esempio, utilizzato per raccogliere stimoli audiovisivi allo scopo di stabilire quali relazioni intercorrano tra emozioni reali ed emozioni posate o recitate, rilevando come le emozioni posate vengano percepite dai soggetti a cui gli stimoli venivano presentati come più intensi o estremi rispetto alle emozioni reali, soprattutto nel caso di emozioni negative.¹⁷⁵ Analoghi risultati sono emersi in un altro studio con lo stesso metodo (Shahid, Kraemer, Swerts, 2008) che mirava ad analizzare la differenza tra emozioni reali ed emozioni posate da una duplice prospettiva (emittente/ricevente) confrontando parlanti sud asiatici (pakistani) e parlanti caucasici (olandesi) senza sostanziali differenze a livello cross-culturale.

In letteratura il metodo si ritrova in diverse varianti e con differenti gradi di riuscita. Proprio in riferimento al grado di efficacia di tale metodo e di altri metodi simili si sono pronunciati in molti, tra cui Gerrards-Hesse *et al.* (1994), Westermann *et al.* (1996), ma anche Clark (1983). Secondo quest'ultimo, ben un terzo, se non addirittura la metà dei partecipanti, manifestano poco o nessun cambiamento di umore in risposta a tale metodo. Dello stesso avviso è anche Kenealy (1986: 331), che in uno studio condotto a partire dalle risultanze sperimentali della Velten MIP su 46 esperimenti, giunge alla conclusione che:

“the findings relating to the Velten procedure [...] are inconsistent and equivocal”.

Larsen, Sinnett (1991) giungono invece alla conclusione che le MIPs sono più efficaci se somministrate con l'inganno¹⁷⁶.

6.2.2 Emozioni “autentiche” o naturali

Un'altra strada molto praticata nello studio delle emozioni espresse per mezzo della voce prevede l'utilizzo di parlato emotivo “reale” o “autentico” che, come detto, risulta di grande interesse per le opportunità e per le possibilità che questo tipo di materiale offre in applicazioni commerciali *real-time*.

¹⁷⁴ Le affermazioni, dell'ordine di ca. 60 frasi, vengono presentate su singole schede (*cards*).

¹⁷⁵ Va comunque rilevata una limitazione segnalata dagli stessi autori che riguarda la presentazione di stimoli solo visivi ai soggetti coinvolti nella somministrazione dei test percettivi.

¹⁷⁶ Sull'uso dell'inganno per la raccolta di materiale in generale, e di materiale emotivamente caratterizzato nello specifico, sussistono una serie di limitazioni di tipo etico morale a cui si è già accennato più sopra.

Riguardo questo aspetto Campbell (2000) mette però in evidenza come, ad esempio, i database utilizzati in ambito di *speech technology* facciano per lo più uso di parlato prodotto da attori, rischiando di rappresentare in modo falsato le caratteristiche del parlato emotivo reale con conseguenze a danno di coloro che sono i fruitori delle applicazioni create e delle soluzioni tecnologiche che su tali materiali vengono modellate:

“If a computer speech synthesizer were to emulate such speaking-style characteristics successfully, then it may be liable to misrepresent the intentions of its user. For example, if a disabled person for whom speech synthesis is the sole means of verbal expression, was to use the synthesized voice to express genuinely-felt pleasure (or anger), then the listener might be able to ‘hear’ that the voice was only expressing acted pleasure (or anger), and is liable to (mis-)respond accordingly.”¹⁷⁷

Tuttavia, seguendo il consiglio di Campbell (2000), sebbene l'utilizzo di emozioni raccolte in contesti o situazioni di vita reale sia da preferire, in quanto caratterizzate da una elevatissima validità ecologica, come fanno notare Scherer, Johnstone, Klassmeyer (2003: 436), anche per questa tipologia di parlato emotivo sussistono una serie di problematiche connesse: l'utilizzo di parlato emotivo appartenente a questa tipologia viene per lo più criticato per il fatto che esso avviene in situazioni di vita reale dove diventa pressoché impossibile definire con esattezza quale emozione venga esperita in quel preciso momento dal soggetto che viene registrato; vi sono poi espressioni che possono esprimere emozioni miste; a ciò si aggiunga anche che spesso le produzioni raccolte sono di brevissima durata e di bassissima qualità per l'impossibilità di controllare le modalità di acquisizione rispettando le più basilari tecniche di registrazione, registrazioni ulteriormente affette da eventi provenienti dall'ambiente circostante.

Ciononostante, come osserva Scherer (2003: 232) il ricorso a materiale di questo tipo si è rivelato, in alcuni casi, estremamente utile, se non addirittura ideale:

“The use of naturally occurring voice changes in emotionally charged situations seems the ideal research paradigm.”

Amir *et al.* (2003), ad esempio, hanno fatto ricorso a parlato emotivo spontaneo ricavato da una serie di *talk show* che hanno avuto luogo a Tel Aviv in Israele, in un contesto in cui la situazione politica risulta assai complessa e in cui i politici tendono ad essere fortemente coinvolti a livello emotivo per evidenti divergenze di opinioni. Dalla significativa

¹⁷⁷ Cfr. Campbell, N. (2000), Databases of emotional speech, in *SpeechEmotion-2000*, pp. 34-35. Sui database di parlato emotivo si ritornerà più avanti in separata sede.

mole di materiale raccolto è stata poi estrapolata per 8 soggetti una selezione di 12 frasi connotate come *angry speech* e 12 come *natural speech*. Su tale materiale sono state poi condotte analisi acustiche come anche analisi sul contenuto testuale delle frasi scelte. Senza scendere in ulteriori dettagli, è qui sufficiente rilevare come, a causa della complessità delle emozioni espresse, vi sia maggiore variabilità nel parlato naturale (nell'accezione di spontaneo) che non in quello posato o recitato.

Altri ancora hanno fatto uso di parlato raccolto in situazioni reali da *call center* di emergenza sanitaria in Francia (Vidrascu *et al.*, 2005).

Nonostante le riserve implicite in questa tipologia di materiale, non vi è alcun dubbio circa il suo elevatissimo grado di validità ecologica che, in determinate situazioni (come ad esempio quelle che prevedono sessioni di addestramento di sistemi di varia natura), rende questo tipo di parlato una fonte di dati estremamente utile.¹⁷⁸

6.2.3 Emozioni simulate o “posate”

Solitamente, quando si parla di emozioni simulate, o ancora posate o recitate, la mente di chiunque chiama inevitabilmente in causa gli attori. Come si vedrà più avanti non è sempre così.

L'utilizzo di attori per la raccolta di parlato emotivamente caratterizzato risale a quelli che vengono conosciuti come *Fairbanks studies* sul finire degli anni '30, considerati da molti come i primi studi sistematici di tipo sperimentale sul parlato emotivo.¹⁷⁹

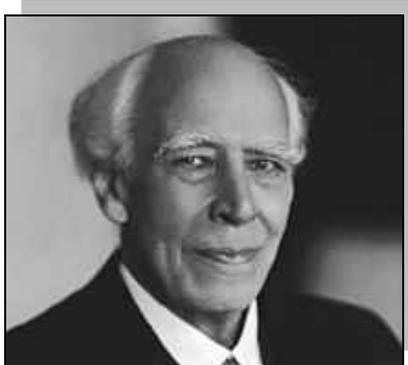


Figura 31: Konstantin Sergeevič Stanislavskij (1863-1938)¹⁸⁰

¹⁷⁸ Per una più ampia rassegna di quelle che sono le metodologie, come anche gli studi, che si sono occupati della raccolta e dell'utilizzo di materiale di questo tipo si veda Scherer (2004).

¹⁷⁹ Vedi Fairbanks, G., Hoaglin, L. W. (1941), An experimental study of the durational characteristics of the voice during the expression of emotion, *Speech Monograph*, 8: 85-91 e Fairbanks, G., Pronovost, W. (1939), An experimental study of the pitch characteristics of the voice during the expression of emotion, *Speech Monograph*, 6: 87-104, *op. cit.* in Schröder (2004), pag. 43.

¹⁸⁰ Fonte: <http://images.encarta.msn.com/xrefmedia/sharemed/targets/images/pho/t039/T039984A.jsm>

In questo caso il metodo di cui solitamente ci si avvale è quello “Stanislavskij”. Si tratta di una tecnica di insegnamento della recitazione messa a punto da Kostantin Sergeevič Stanislavskij nei primi anni del '900.

Il metodo, che nasce da alcune sue puntigliose annotazioni relative alle proprie esperienze di vita come scrittore, attore e regista, si basa sull'approfondimento psicologico del personaggio e sulla ricerca di affinità tra il mondo interiore del personaggio e quello dell'attore. Si basa sulla esternazione delle emozioni interiori attraverso la loro interpretazione e rielaborazione a livello intimo e a livello di esperienze del vissuto personale dell'attore stesso. Per ottenere la credibilità scenica il maestro Stanislavskij proponeva esercizi volti a riproporre le emozioni da provare sulla scena, e quindi da trasmettere alla platea degli uditori, andando ad analizzare in modo profondo gli atteggiamenti non verbali e il sottotesto del messaggio da trasmettere.

Si tratta, in definitiva, di quello che può essere considerato un processo di auto-induzione di stati emotivi a partire dalla propria esperienza di vita, un metodo oggi insegnato nelle più prestigiose scuole di recitazione di tutto il mondo.¹⁸¹

L'utilizzo di questo metodo per la raccolta di parlato emotivamente caratterizzato a fini di ricerca pone, però, una serie di questioni. Tra quelli che possono essere definiti vantaggi vi sono sicuramente la possibilità di ottenere materiale emotivo anche molto “intenso” che ricade positivamente sulla possibilità che le produzioni prodotte vengano correttamente identificate o riconosciute da gruppi di ascoltatori in prove di tipo percettivo. Questa “intensità” delle produzioni raccolte si scontra, tuttavia, con quella che è rappresentata dalla realtà e dalla vita quotidiana: in che misura, in condizioni normali come quelle della vita di tutti i giorni, si assiste a realizzazioni così intense? A questo proposito Scherer (2003: 232) sottolinea come si tratti di un rischio noto affermando appunto come:

“it cannot be excluded that actors overemphasize relatively obvious cues and miss more subtle ones [...]”

L'attore, nel rivivere e riprodurre determinate emozioni, tende forse a rappresentare quella che dovrebbe essere la realtà in modo troppo esagerato o stereotipato: questo porterebbe sicuramente ad una maggiore riconoscibilità della produzione emotiva, ma indurrebbe allo stesso tempo a considerare innaturale o poco realistica l'emozione codificata. Un altro aspetto a sfavore di questa tecnica in contesti scientifici quali quelli previsti dalla

¹⁸¹ Quelli che sono considerati i suoi insegnamenti sono stati raccolti da Stanislavskij stesso in *Il lavoro dell'attore su sé stesso* pubblicato nel 1938 e in una pubblicazione dal titolo *Il lavoro dell'attore sul personaggio* apparso postumo nel 1957.

ricerca, è dato dal fatto che non vi è assolutamente alcun controllo su quelle che sono le esperienze (soggettive) richiamate dal singolo attore nella realizzazione delle emozioni richieste: ciò causa inevitabili ripercussioni su quello che è il significato che una determinata emozione assume da un soggetto all'altro e che è strettamente legato alla situazione che l'ha generata. Infine, se da un lato il ricorso a questa tecnica consente di ridurre al minimo il dispendio di tempo per la raccolta del corpus, in quanto si tratta di produzioni in un certo senso mirate ai fini della ricerca che si intende condurre, dall'altro si pone il problema del reperimento di attori professionisti che si prestino a simili operazioni e che raramente, fatte salve alcune eccezioni, prestano la loro opera gratuitamente.

La Tabella 10 di seguito proposta riassume quelli che possono essere considerati gli elementi a vantaggio o a svantaggio derivanti dall'uso di questa particolare tecnica di raccolta che, per quelli che sono i suoi elementi caratterizzanti, prevedono un esclusivo utilizzo di attori.

Pro	Contro
Emozioni molto "intense"	Variabilità delle emozioni da un attore all'altro in base alle situazioni richiamate
Emozioni altamente riconoscibili	Espressioni stereotipate e posate/simulate
Poche registrazioni necessarie	Scarsa rappresentazione della realtà
Poco dispendio di tempo	Disponibilità degli attori ed eventuale retribuzione per l'opera prestata

Tabella 10: Tabella riassuntiva dei vantaggi e degli svantaggi relativi al metodo Stanislavskij

Nonostante i vantaggi e i limiti appena esposti, si tratta della strada percorsa dalla maggior parte degli studiosi in questo ambito: sono tanti, infatti, coloro che fino ad oggi hanno fatto ricorso a questa tecnica. Ad esempio Burkhardt *et al.* (2005) hanno fatto uso di questa metodologia per la creazione di un database di parlato emotivo recitato in lingua tedesca contenente 10 frasi prodotte in sei stati emotivi target da 10 attori, fornendo agli attori coinvolti semplici istruzioni orali e facendo esplicito affidamento sulle loro abilità ad auto-indursi determinate emozioni:

*"We relied on the performers' ability of self-induction by remembering a situation when the desired emotion had been felt strongly, which is known as the Stanislavski method."*¹⁸²

In modo analogo ha operato Iadarola (2009) nella raccolta di un database di parlato emotivo per l'italiano in cui i soggetti coinvolti erano tutti attori e per i quali l'autore si è:

¹⁸² Cfr. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, Walter F., Weiss, B. (2005), A database of German emotional speech, in *INTERSPEECH-2005*, pag. 1517.

“[...] premurato di verificare i loro trascorsi professionali, in particolar modo la loro conoscenza del famoso metodo Stanislavskji [...].”¹⁸³

sottolineando e giustificando preventivamente la propria scelta, rispondendo indirettamente alle critiche di quanti si oppongono all'utilizzo di questa metodologia (di cui si è più sopra discusso), nei seguenti termini:

“E' proprio la ragion d'essere dell'attore quella di riprodurre le emozioni della vita reale, e una delle prime cose che un attore impara è di non essere impostato, ma 'naturale'.”¹⁸⁴

Va tuttavia tenuto presente che gli studi basati esclusivamente su quelle che sono le linee guida imposte dal metodo Stanislavskji sono di numero decisamente inferiore di quelli che, per non incorrere nelle critiche rivolte a questa tecnica, hanno fatto ricorso ad attori con accorgimenti e strategie metodologiche diverse volte proprio a superare, o quanto meno a ridurre, alcune significative limitazioni.¹⁸⁵ Dagli studi censiti da Juslin, Laukka (2003) è infatti emerso che nell'87% dei casi sono state utilizzate emozioni posate o simulate da attori, mentre in Ververidis, Kotropulos (2006) delle 64 risorse di parlato emotivo riportate ben 36 risultano contenere parlato simulato.

A questo proposito Enos, Hirschberg (2006) propongono due particolari approcci per l'elicitazione di parlato emotivo con attori, proprio per far fronte alle difficoltà che pone l'utilizzo del parlato recitato nell'ambito delle ricerche sulle emozioni. Il primo approccio ripercorre quelle che sono le strade e i metodi seguiti in precedenti studi da altri ricercatori, attuando un processo di *reverse-engineering* dell'emozione desiderata, fornendo cioè all'attore una destrutturazione dell'emozione nelle componenti che a quest'ultimo sono più fruibili e che gli vengono fornite nei termini riportati in Figura 32.

In questo caso si fa ricorso ad una versione particolarmente elaborata del paradigma degli scenari, su cui si ritornerà più avanti, in cui i cinque elementi (vedi Figura 32) vengono intesi come un'equazione che produce come risultato il comportamento umano, emozione inclusa. In questo contesto, seguendo le linee guida degli autori, il ricercatore, coadiuvato da un bravo regista, non fa altro che regolare i parametri di detta equazione ai fini del

¹⁸³ Cfr. Iadarola (2009), pag. 299.

¹⁸⁴ *Ibidem.*

¹⁸⁵ Il problema non è ovviamente legato al metodo in quanto tale, ma al suo utilizzo in questo specifico ambito. Uno degli elementi che suscita maggiore critica è dato dal fatto che il più delle volte all'attore venga semplicemente chiesto di pronunciare un testo “X” con una emozione “Y” senza ulteriori istruzioni aggiuntive, lasciando appunto troppo spazio alla “interpretazione” soggettiva del singolo.

raggiungimento del risultato desiderato. Compito dell'attore sarà infine quello di pronunciare un testo in modo da incidere sull'obiettivo, mantenendo l'attenzione sull'azione.

1. A description of the character.
2. A clear description of the situation, including obstacles to the goal, and what is at stake.
3. A (single) goal for the actor.
4. A text that is meant to accomplish the goal, regardless of whether the specific lexical content embodies that goal.
5. Description(s) of additional character(s) involved, if any.

Figura 32: Elementi forniti all'attore nel processo di *reverse-engineering* dell'emozione desiderata¹⁸⁶

Affinché quanto descritto abbia validità, il tutto deve ovviamente avvenire:

“without mention to the actor of the desired emotion.”¹⁸⁷

ovvero, senza fornire all'attore informazione alcuna sull'emozione che con queste o quelle istruzioni fornite si vuole raccogliere.

Nel secondo, invece, si fa un uso diretto di quelli che sono i meccanismi, le abilità e i materiali più familiari agli attori attraverso quello che gli autori definiscono un *well-crafted play* (ovvero un pezzo artisticamente più impegnativo, ma allo stesso tempo molto familiare all'attore)¹⁸⁸, seguendo i cinque punti di cui all'approccio sopra descritto. Così facendo Enos, Hirschberg (2006: 9) sottolineano come:

“This approach avoids conveying specific emotional labels to the actor in favor of providing goals and contexts in which to play out certain actions.”

riducendo perciò al minimo il problema di fornire ai soggetti specifiche etichette emotive che possono a priori condizionare e influenzare le produzioni emotive realizzate.

Concludendo, si può tuttavia dire che il ricorso alle emozioni risulta in un certo qual modo giustificato dal fatto che:

“[...] emotion portrayals reflect sociocultural norms or expectations more than the psychophysiological effects on the voice as they occur under natural conditions.”¹⁸⁹

Si tratta di un dato di fatto che porta Scherer ad affermare:

¹⁸⁶ Cfr. Enos, Hirschberg (2006), pag. 8.

¹⁸⁷ *Ibidem.*

¹⁸⁸ Gli autori propongono come esempio un passaggio tratto da Romeo e Giulietta di Shakespeare.

¹⁸⁹ Cfr. Scherer (2003), pag. 232.

*“it can be argued that all publicly observable expressions are to some extent “portrayals”.*¹⁹⁰

In fondo, nella vita di tutti i giorni, ciascuno di noi incarna un po' quella che può essere considerata la figura dell'attore.

6.2.4 Gli scenari: una soluzione “ibrida”

Si è più sopra fatto accenno all'utilizzo di scenari, o *scenario approach*, per la raccolta di parlato emotivo.¹⁹¹ Con il termine ‘scenario’ ci si riferisce ad una descrizione sintetica di un evento o di una serie di azioni e di eventi, la cui organizzazione e/o composizione segue quelle che sono le tecniche narrative. Come già evidenziato per molte altre tecniche, anche questo tipo di approccio si presenta in diverse varianti.

Nella variante utilizzata da Anolli *et al.* (2008a, 2008b) e da Anolli, Ciceri (1992), in cui gli scenari sono stati somministrati a soggetti *naïf*, si prevede la presenza e la definizione di due elementi fondamentali: a) una situazione che fornisce le condizioni caratteristiche della specifica emozione; b) una serie di risposte standard alla situazione. A partire da questi due elementi in ciascun testo sono state inserite in modo sistematico informazioni:

- “stimolo pertinenti”, ovvero informazioni su stimoli esterni in grado di attivare una data emozione, stimoli allo stesso tempo coerenti con l'emozione in questione;
- “risposta pertinenti”, ovvero una serie di informazioni riguardo alle risposte emotive più consone al contesto e comprendente risposte verbali, esclamazioni, reazioni motorie ecc.;
- “di significato”, ovvero informazioni che definiscono il significato dello stimolo emotigeno e delle risposte da esso derivanti con riferimento all'*appraisal*.¹⁹²

¹⁹⁰ *Ibidem.*

¹⁹¹ L'approccio degli scenari potrebbe essere considerato come metodo di induzione di emozioni e sarebbe, quindi, dovuto essere trattato nel paragrafo dedicato al parlato emotivo indotto. Tuttavia, come si vedrà nel prosieguo, questo approccio lascia un certo grado di libertà ai soggetti coinvolti nelle operazioni di raccolta di parlato emotivo, e più che “indurre” determinate emozioni si parla in questo caso di “elicitare” le emozioni facendole, in un certo qual modo, emergere o affiorare con un certo grado di naturalezza, consentendo, allo stesso tempo, di tenere sotto controllo una serie di variabili a cui si accennerà più avanti.

¹⁹² *Cfr.* Anolli *et al.* (2008a), pag. 10. Gli autori seguono le proposte di precedenti studiosi tra cui De Sousa (1987) e Rosenthal *et al.* (1979), utilizzando anche una versione adattata del “paradigma del contenuto standard” di Davitz (1964) che prevede l'inserimento di una frase dal contenuto standard in tutti gli scenari proposti. Nella variante presentata, per la lingua italiana la frase era “non è possibile, non ora” (*cfr.* Anolli *et al.* 2008a, pag. 8). Per gli scenari utilizzati si rimanda ad Anolli *et al.* (2008b) e Anolli, Ciceri (1992).

Date queste premesse, gli scenari creati sono stati fatti leggere ad alta voce ad un gruppo di soggetti.¹⁹³ Successivamente, in un secondo momento, la frase dal contenuto standard è stata estrapolata ed isolata dal contesto per le successive fasi della ricerca.



Figura 33: Rappresentazione schematica del costrutto di uno scenario o *frame story* nei termini proposti da Anolli *et al.* (2008a, 2008b) e Anolli, Ciceri (1992)

Secondo gli autori di questo approccio vi sono una serie di buone ragioni che ne favorirebbero l'adozione rispetto ad altre metodologie, ragioni che vengono così sintetizzate:

“(a) closeness to real life situations (the stories follow a narrative thinking format in which the mind is engaged in sequential, action-oriented, context-bound, detail-driven thought, attaining a high degree of ecological validity; Bruner, 1991); (b) attention catching ability (a well-designed story has a progressively unfolding plot capable of focusing and sustaining the reader’s attention), (c) in the case of reading aloud, there is concurrence of emotion activation and its vocal expression (any temporal delay between these two processes can be a source of error due to systematic biases in recall [Levenson, 1988]; the story technique enables emotion, however fleeting, to be captured as it occurs—when reading a story out loud, individuals inevitably express with their voice what they are feeling inside), (d) possibility to create matching story formats (the different stories, each focused on a specific emotion, can be standardized for length, order and arrangement of events, context sensitivity, and so on), and (e) possibility to elicit emotion while incurring a relatively low level of demand vis-à-

¹⁹³ Nello specifico sono stati utilizzati 29 studenti cinesi e 19 studenti italiani che hanno partecipato alla fase di *encoding* del materiale emotivo in qualità di attori-lettori ai quali era stato detto di partecipare, in fase di reclutamento, ad *“uno studio sul gradimento qualitativo di determinate caratteristiche della voce umana commissionato da una radio regionale”* (cfr. Anolli *et al.* 2008a, pag. 8 e segg.). Prima della somministrazione, gli scenari sono stati preventivamente valutati con un gruppo di giudici *“al fine di garantire il riconoscimento univoco dell’emozione, la validità interna e di costruito delle storie...”* (cfr. Anolli *et al.* 2008a, pag. 10).

*vis time, setting, minimization of intrusiveness, ease of use, and so on.*¹⁹⁴

Nell'altra variante di questo approccio, utilizzato sino ad oggi da un numero maggiore di studiosi, è invece prevista la somministrazione, in questo caso, ad attori professionisti, di scenari o situazioni molto più sintetici, sulla base dei quali vengono fornite agli attori coinvolti, specifiche istruzioni.¹⁹⁵ Nel caso di Scherer *et al.* (1991), gli scenari sono molto sintetici, come ad esempio quelli utilizzati per *anger*:

*"1) The director is again late for rehearsal and we have to work until very late at night. Once again I have to cancel a date. 2) I have sublet my apartment for a period of several months. Upon my return my place is in a real mess and the person did not keep to a single agreement."*¹⁹⁶

In questo caso agli attori è stato chiesto di leggere gli scenari e di immaginare di vivere o di esperire la situazione descritta e successivamente di pronunciare una frase standard¹⁹⁷ nello stesso modo in cui l'avrebbero pronunciata se si fossero trovati in quella situazione.

Il concetto di "ibrido", a cui si è fatto qui riferimento, consiste nella possibilità di indurre e allo stesso tempo di far simulare (entro certi limiti) su una base comune (data dagli scenari) le emozioni desiderate utilizzando sia soggetti *naïf* che *attori*. Tale approccio consentirebbe di aggirare alcune delle limitazioni di cui ai precedenti paragrafi, sebbene, come già evidenziato per le precedenti modalità di raccolta di parlato emotivo, anche in questo caso non mancano vantaggi e svantaggi che vengono riportati nella Tabella 11:

Pro	Contro
Controllo della situazione elicitante e delle emozioni	Gli scenari vanno creati o adattati allo scopo
Facilità di attuazione	Molti soggetti da registrare
Abbondanza di materiale	Richiede molto tempo

Tabella 11: Tabella riassuntiva dei vantaggi e degli svantaggi relativi ad una soluzione di raccolta "ibrida" a cavallo tra emozioni simulate ed emozioni indotte

¹⁹⁴ Cfr. Anolli *et al.* (2008b), pag. 569.

¹⁹⁵ Considerando, quindi, quest'altro aspetto, l'argomento qui trattato avrebbe dovuto trovare spazio nel paragrafo dedicato alle emozioni simulate o "posate".

¹⁹⁶ Cfr. Scherer *et al.* (1991), pag. 147. Gli scenari proposti dagli autori sono stati ricavati da studi di tipo interculturale precedentemente condotti in cinque continenti sull'esperienza delle emozioni in cui sono state raccolte situazioni rappresentative dell'elicitazione delle emozioni su un campione di oltre 3.000 intervistati (cfr. Wallbott, Scherer, 1986).

¹⁹⁷ In questo caso sono state usate "*speech-like but meaningless sentence(s)*" create da un fonetista sulla base di elementi fonemici di varie lingue europee: "*Hat sundig pron you venzy.*" e "*Fee gott laish jonkill gosterr.*" (*ibidem*, pag. 126). La scelta di questo tipo di frasi era stata fatta in vista di un successivo studio a livello cross-culturale.

6.3 Scelta del materiale linguistico

La scelta del materiale linguistico, o del tipo di parlato, da utilizzare nelle ricerche sul parlato emotivo rappresenta, da questo punto di vista, un elemento cruciale che merita la dovuta considerazione.

Strettamente connessa al materiale linguistico scelto risulta infatti la possibilità di effettuare determinate analisi di tipo acustico, o ancora, di tipo percettivo.

Sebbene le possibilità di scelta siano, da questo punto di vista, pressoché infinite (considerando le caratteristiche di ricorsività della lingua), queste sono sostanzialmente riconducibili a frasi (come in Anolli, Ciceri, 1992; Walbott, Scherer, 1986), frasi non-sense (Scherer *et al.* 2001), ma anche parole isolate (Piôt, 1999), cifre numeriche, sequenze ‘VCV (come ad esempio /’aba, ‘ava/ in Magno Caldognetto *et al.*, 2005), vocali isolate (Magno Caldognetto, Kori, 1986) o ancora affect bursts (Schröder, 2003).

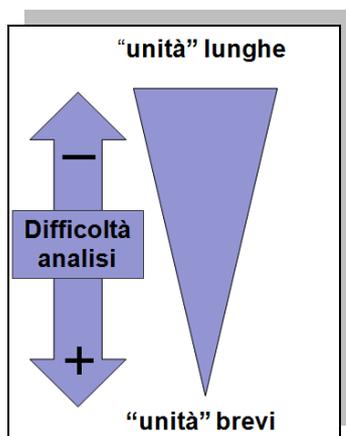


Figura 34: Rappresentazione schematica del livello di difficoltà connesso alle unità di parlato scelte

Come è facile intuire, nell’elenco delle possibilità appena riportato, a cui si aggiungono le caratteristiche sulla tipologia di parlato individuate da Fagyal (1995) e schematizzate in Figura 30, non solo la quantità di materiale utile diminuisce a seconda della tipologia di materiale utilizzato, ma di contro aumentano le difficoltà delle analisi che sulle produzioni raccolte si intendono effettuare.

6.4 Scelta dei parlanti: parlanti naif vs. attori

Molte delle differenze legate a questo aspetto sono già state ampiamente trattate, evidenziando vantaggi e svantaggi nell’utilizzo delle metodologie di raccolta di parlato emotivo che, a seconda della scelta e degli obiettivi, producono parlato emotivo “reale” o “autentico” o ancora recitato o “posato”. In base alla metodologia di raccolta adottata per la raccolta di produzioni vocali delle emozioni, ci si limiterà qui soltanto a dire che per

perseguire tali obiettivi si aggiunge, come detto, la possibilità di utilizzare parlanti in qualità di soggetti *naïf* o parlanti in qualità di *attori* e questo indipendentemente dalla metodologia di raccolta utilizzata.

Si è già accennato, ad esempio, allo studio di Wilting, Kraemer, Swerts (2006) che ha messo in evidenza le differenze nell'uso di attori nel caso si vogliano studiare emozioni autentiche piuttosto che posate e di come vi sia, da parte dei soggetti a cui gli stimoli sono stati presentati, una maggiore percezione di emozioni più intense prodotte da soggetti attori, soprattutto per emozioni negative. Questo rappresenta, di fatto, uno dei maggiori "rischi", se così si può definire, a cui si può andare in contro qualora si decidesse di utilizzare soggetti attori: il rischio è principalmente dato dalla tendenza all'esagerazione (*overacting*) di determinati fenomeni. Nel caso in cui ci si trovi in situazioni che richiedono un *setting* sperimentale altamente controllato, il ricorso a soggetti *attori* rimane, in definitiva, l'unica soluzione che permette una assoluta supervisione del *task*. Sarà in questo caso il protocollo di raccolta stilato a ridurre o a limitare fenomeni come quelli dell'*overacting*.

Chiusa questa breve parentesi, al ricercatore interessato alla raccolta di parlato emotivo si presentano diverse possibili opzioni che gli consentono di combinare tra loro gli "ingredienti" della pietanza che comporrà il protocollo di raccolta da somministrare ai soggetti da registrare.

6.5 Etichette verbali emozionali

Nello studio delle emozioni vocali anche le "semplici" etichette verbali, che compongono il lessico emotivo di ciascun parlante (come ad esempio gioia, tristezza e collera), hanno la loro importanza.

Come già parzialmente accennato, le emozioni presentano rilevanti variazioni culturali sia nella loro natura sia nelle loro modalità espressive. Accanto a questo aspetto va parimenti tenuto in conto che ogni cultura ha elaborato e possiede un proprio lessico emotivo, in funzione del quale gli individui che ad essa appartengono attribuiscono un nome alle proprie esperienze emotive, in base alle quali sono in grado di comunicarle e di riconoscerle negli altri individui della propria comunità.

Uno degli aspetti spesso trascurato durante la raccolta di parlato emotivo riguarda il fatto che una singola etichetta emozionale possa rimandare o riferirsi a più di uno stato affettivo: non esiste la collera, ma la collera fredda e la collera calda.¹⁹⁸

¹⁹⁸ In modo del tutto analogo, nella vita di tutti i giorni, non è raro assistere all'espressione di emozioni "miste".

Nel caso vengano condotte ricerche all'interno di una singola lingua le difficoltà maggiori sono legate soprattutto al fatto che soggetti diversi possono attribuire diverso significato emotivo a specifiche etichette verbali, come nel caso già richiamato. O ancora, come evidenziano Scherer *et al.* (2001: 79), che soggetti diversi possono usare situazioni elicитanti diverse per una stessa emozione.

La questione si complica ulteriormente nel caso in cui la ricerca o lo studio preveda che queste debbano essere tradotte in un'altra lingua, come nel caso di studi di tipo cross-linguistico. Specifici termini emotivi presenti in una lingua, ad esempio nell'italiano, possono anche non avere un equivalente semantico e lessicale in un'altra lingua o, ancora più frequentemente, in un'altra cultura. Prima di affrontare tali questioni più da vicino, si dia uno sguardo alla Tabella 12 che costituisce il punto di partenza nel caso si vogliano affrontare ricerche sulle emozioni a livello cross-linguistico:

Italiano	Français	English	Deutsch
rabbia (calda)	colère (chaude)	anger (hot)	Ärger (warmes)
gioia	joie	joy	Freude
paura	peur	fear	Angst
disgusto	dégoût	disgust	Ekel
tristezza	tristesse	sadness	Traurigkeit
sorpresa	surprise	surprise	Überraschung
neutro	neutre	neutral	Neutral

Tabella 12: Etichette verbali relative a sei stati affettivi in quattro lingue europee¹⁹⁹

Da quanto riportato in Tabella 12 si emerge una corrispondenza uno a uno dei termini riportati in appendice nel volume curato da Scherer (*op. cit.*). Tuttavia, la corrispondenza non è mai così scontata e, alle problematiche già individuate all'interno di ogni singola lingua, si aggiungono qui ulteriori difficoltà.

Come evidenziato da Averill (2004: 578), le etichette verbali differiscono da lingua a lingua e non sempre il rapporto tra i termini è univoco. Dove una lingua, espressione di una cultura, e in quanto tale interpretazione della realtà, contempla un solo termine per uno o più stati affettivi, in un'altra viene operata una netta distinzione dei termini, attribuendo determinate etichette emozionali a determinati stati affettivi: ad es., contrariamente all'italiano che non distingue ulteriormente il concetto di 'tristezza', identificato da una sola etichetta, il tedesco prevede un'etichetta *Traurigkeit* per la tristezza comunemente intesa, mentre si riferisce a *Trauer* nel caso di tristezza legata ad eventi luttuosi. A ciò si aggiunga, inoltre, che

¹⁹⁹ Tabella adattata da Scherer, K. R. (1988), Labels describing affective states in five major languages, in, *Facets of emotion: Recent research*, K. R. Scherer (Ed.), Hillsdale, NJ: Erlbaum, pp. 241-243.

una stessa etichetta emozionale può avere una sola valenza in una lingua (senza distinzione tra stato positivo o negativo), mentre può avere tale distinzione in un'altra lingua. Härtel, Härtel (2005: 685-686) fanno infatti notare come, ad esempio,

*“the word 'surprise' in English does not necessarily have a positive or negative connotation, whereas the German word 'Überraschung' probably has more of a positive than a negative connotation.”*²⁰⁰

Altro esempio potrebbe essere quello di ‘gioia/felicità’ (*joy/happiness*): nella cultura anglosassone il termine *happy*, molto più frequente di *joyful*, non corrisponde pienamente alla categoria italiana ‘felice’.

Insomma, ribadendo ancora una volta quanto più volte detto, le emozioni sono strettamente legate al sistema culturale all'interno del quale esse si manifestano. È tale sistema che fornisce ai suoi appartenenti parametri cognitivi e specifici modelli per interpretare la realtà e per rispondere ad essa in modo adeguato ed il non tenerne conto potrebbe irrimediabilmente compromettere gli esiti della ricerca intrapresa.

6.6 Corpora o risorse di parlato emotivo

I *corpora*, o database in generale, sono raccolte più o meno ampie di parlato, di produzioni verbali (scritte o orali) raccolte a vario titolo, che rientrano in quelle che comunemente vengono identificate con il termine di “risorse linguistiche”. In questo contesto con il termine di *corpora* ci si riferisce esclusivamente a risorse linguistiche di parlato emotivo tralasciando, volutamente, qualsiasi altra tipologia di risorsa.

Grazie all'oramai inarrestabile innovazione tecnologica si assiste, allo stato attuale, ad un fiorente mercato attorno a tali raccolte, che vengono commissionate e progettate da vari soggetti o enti per gli scopi più svariati, soprattutto per applicazioni in ambito commerciale. Esistono enti o istituzioni come la *European Language Resources Association* (ELRA) la cui missione è quella di promuovere le risorse linguistiche nel settore dell'*Human Language Technology* (HLT) e di valutare le tecnologie di ingegneria della lingua, per conto della quale opera un'altra agenzia denominata ELDA (*Evaluations and Language resources Distribution Agency*) che si occupa, nello specifico, della valutazione e della distribuzione di risorse

²⁰⁰ Per una rassegna più completa ed esauriente si rimanda il lettore all'opera di Pavlenko, A. (2005), *Emotions and Multilingualism*, New York: Cambridge University Press, in cui tali aspetti vengono chiariti e messi in relazione con l'ausilio di numerosi esempi.

linguistiche.²⁰¹ Per quello che in questa sede può interessare, è assai singolare il fatto che, all'interno del catalogo delle risorse accessibili dal sito web dell'agenzia, una ricerca effettuata con la parola chiave “*emotion*” abbia prodotto come unico risultato un rimando a sole tre raccolte di uno stesso progetto.²⁰²

Occorre a questo punto richiamare una distinzione operata da Douglas-Cowie *et al.* (2003) che forse può contribuire a fare luce sull'apparente carenza di risorse di parlato emotivo nei canali “ufficiali” incaricate della distribuzione di tali risorse. Più che assistere alla presenza di veri e propri *database*, nella letteratura sulle emozioni vocali si assiste alla presenza di *dataset*, ovvero di raccolte di dimensioni relativamente ridotte, generalmente costituite per studiare singoli aspetti del parlato emotivo.

Come fanno notare Douglas-Cowie *et al.* (2003: 33) la ricerca sul parlato e sulle emozioni,

“is moving from a period of exploratory research into one where there is a prospect of substantial applications, notably in human-computer interaction.”

Il fatto appena riportato non è poi del tutto banale in quanto, come già detto, le raccolte di parlato emotivo attualmente esistenti sono il frutto di vari esperimenti e di varie ricerche sul parlato emotivo condotte negli anni. Quello delle risorse linguistiche è un tema assai caro alla comunità dello *speech* dove non mancano certo i tentativi di creare simili raccolte. È proprio sulla base delle esperienze, delle difficoltà e delle scelte operate negli studi che negli anni si sono succeduti che Douglas-Cowie *et al.* (2003) mettono in luce una serie di caratteristiche che le nuove generazioni di *database* dovranno tenere in considerazione:

1. *Scopo*. Un database dovrebbe considerare e includere il maggior numero di variazioni possibili, come ad esempio numero di parlanti, lingua parlata, sesso, emozioni considerate ecc., che ne consentano un utilizzo adeguato, generalizzato ed ampio. La sua progettazione e creazione dovrebbe, inoltre, tener conto dei possibili ambiti di utilizzo.
2. *Naturalzza*. Esistono diverse strategie di raccolta del parlato emotivo, strategie che danno origine a produzioni naturali, semi-naturali e simulate o posate. Sulla

²⁰¹ Per maggiori dettagli si rimanda a <http://www.elra.info/>.

²⁰² Il progetto è il VERBMOBIL II, un progetto a lungo termine del Ministero Federale dell'Educazione e della Ricerca tedesco (*Bundesministerium für Bildung und Forschung – BMBF*) ed ha lo scopo di fornire alla Germania una posizione di punta nelle tecnologie del linguaggio e alle sue applicazioni economiche grazie ad una cooperazione e concentrazione del maggior numero possibile di specialisti provenienti dall'industria e dalla scienza.

scelta delle tipologia di parlato più idonea prevalgono, in ogni caso, gli obiettivi della ricerca che stabilisce quale di esse sia la più adeguata.

3. *Contesto*. L'ascoltatore fa generalmente riferimento al contesto per comprendere il significato emotivo di determinate caratteristiche vocali. Quattro sono i tipi di contesto identificati dagli autori:
 - a. *semantico*: nel parlato emotivo ci si avvale talvolta di parole emotivamente marcate stabilendo una potenziale interazione tra contenuto e segnale vocale;
 - b. *strutturale*: molti eventi o segnali emotivi vengono codificati in base a strutture sintattiche (accenti, *patterns* intonativi ecc.) o in base a variazioni di stile attraverso le caratteristiche strutturali delle frasi (lunghe o corte, ripetizioni, interruzioni ecc.);
 - c. *intermodale*: non sempre il parlato rappresenta una fonte di informazione *stand-alone* (a se stante), ma funge spesso da supplemento ad altre fonti di informazione quale quelli della faccia e dei gesti;
 - d. *temporale*: il parlato naturale veicola determinati indici di variazione lineare e sequenziale nel momento in cui le emozioni si manifestano e svaniscono nel corso del tempo.
4. *Descrittori*. Un *database* richiede tecniche di descrizione del contenuto linguistico ed emozionale da un lato e del parlato dall'altro.
5. *Accessibilità*. Il valore e l'importanza di un *database* aumenta ed è tale solo se esso è accessibile all'intera *speech community*, ciò al fine di ridurre gli sforzi e facilitare i confronti sul medesimo materiale. Due sono sostanzialmente gli aspetti che vincolano l'accessibilità di un *database* e che sono riconducibili a:
 - a. formato, che deve essere standardizzato ed aperto;
 - b. principi etici, che concernono il suo utilizzo e la sua diffusione (maggiormente limitante nel caso di parlato emotivo naturale), a cui si aggiungono anche questioni di *copyright*.

È facile intuire come la generazione di un *database* che rispecchi le caratteristiche individuate dagli autori rappresenti un'impresa assai difficile: solo gli obiettivi e le finalità della specifica ricerca saranno in grado di stabilire e determinare le scelte e le caratteristiche individuate da Douglas-Cowie *et al.* (2003).

Prendendo, invece, in esame la ricognizione di quelle che Ververidis, Kotropoulos (2006) chiamano *emotional speech data collections*²⁰³, o raccolte di parlato emotivo, emerge una verità che nel presente contesto assume una rilevante importanza.²⁰⁴ Andando alla ricerca di una risorsa di parlato emotivo mistilingue che risponda a quelle che sono le tematiche che qui si intende affrontare, si nota l'assoluta assenza di simili risorse tranne qualche eccezione. Le uniche risorse mistilingui riscontrabili contengono stimoli in non più di due lingue: inglese e tedesco nel caso di Batliner *et al.* (2004) e Scherer *et al.* (2002); inglese e sloveno nel caso di Ambrus (2000), e inglese e spagnolo in Gonzalez (1999). Il dato appena riportato non è assolutamente sorprendente se si considera poi che la sezione web dedicata da HUMAINE²⁰⁵ alle risorse disponibili nell'ambito delle emozioni raccoglie, di fatto, 21 risorse per il solo parlato e 13 risorse contenenti materiale audio-visivo ed altro: l'elenco è decisamente meno aggiornato di quello di Ververidis, Kotropoulos (2006).

Ciò che si vuole qui sottolineare non è tanto l'assenza di risorse linguistiche, a cui se ne sono nel frattempo aggiunte altre a cui probabilmente non si è riusciti ad avere accesso, quanto il fatto che le emozioni vocali rappresentino, di per sé, una tematica assai complessa: se nelle singole lingue si incontrano una serie di difficoltà e di problematiche, da quanto è stato fatto qui notare, tali questioni si riflettono e si complicano ulteriormente nelle ricerche di tipo cross-linguistico e cross-culturale.

²⁰³ Ricognizione che comprende e integra la breve lista di 21 *datasets* riportata da Douglas-Cowie *et al.* (2003).

²⁰⁴ È per le ragioni che seguiranno che, in un certo qual modo, la scelta di creare un corpus sperimentale di parlato emotivo mistilingue europeo risulta motivata, ma su questo aspetto si ritornerà in modo più approfondito nel prosieguo.

²⁰⁵ Cfr. <http://emotion-research.net/wiki/Databases> (data consultazione: 10 dicembre 2007). HUMAINE (*Human-Machine Interaction Network on Emotion*) è un Network di eccellenza nato nel 2004 all'interno del 6° Programma Quadro EU con 33 partner provenienti da 14 paesi diversi.

7 Le emozioni a livello cross-linguistico: stato dell'arte

Lo studio delle emozioni espresse attraverso la voce ha subito negli ultimi anni, come più volte ribadito, un rinnovato interesse in diversi ambiti di ricerca e numerosi sono gli studi presenti in letteratura.

Se da un lato si rilevano innumerevoli studi sul parlato emotivo nelle singole lingue, dall'altro gli studi di tipo cross-linguistico e cross-culturale risultano essere assai sparuti. I motivi di questa situazione non sono certamente da attribuire allo scarso interesse che tali ricerche suscitano negli studiosi ma, al contrario, sono prevalentemente da ricondurre a tutta una serie di difficoltà che lo studio delle emozioni vocali impone, difficoltà ulteriormente esacerbate nello studio di tipo cross-linguistico e cross-culturale quale quello che qui si intende intraprendere.

Nel fornire una chiara ed esauriente trattazione di quelli che sono rappresentati dagli studi di tipo cross-linguistico e cross-culturale, si procederà ad una arbitraria suddivisione degli stessi in studi volti al *decoding*, all'*encoding* e in studi volti sia all'uno che all'altro aspetto.

Negli studi di *decoding*, secondo un approccio che viene definito "indiretto" o *listener-centered*, viene esaminata l'abilità di membri appartenenti ad una lingua o ad una cultura a riconoscere correttamente espressioni emotive di una lingua o di una cultura diversa. Di contro, gli studi volti all'*encoding* seguono quello che può essere definito un "approccio diretto" o *speaker-centered*: secondo tale approccio vengono analizzate le risposte emotive di individui, appartenenti a lingue o culture diverse, a specifiche situazioni emotigene in grado di attivare determinate emozioni alla ricerca di un legame tra lo stato emotivo del parlante e le modificazioni indotte nella sua voce o nel suo prodotto vocale.

Nella presente trattazione si è inoltre ritenuto utile considerare anche una categoria di studi in cui siano stati utilizzati sia l'uno che l'altro approccio. Tuttavia, forse per limiti di spazio legati alla pubblicazione dei risultati delle ricerche in alcune riviste o ambiti, le analisi condotte e i risultati conseguiti vengono spesso solo parzialmente riportati e commentati, rendendo a volte difficile stabilire fino a che punto il duplice aspetto dell'*encoding* e del *decoding* siano stati effettivamente affrontati. La maggiore difficoltà, in questi casi, è stata

riscontrata nel riuscire a definire fino a che punto l'aspetto dell'*encoding* sia stato affrontato e trattato in quanto, spesso e volentieri, nelle ricerche condotte ci si riferisce all'*encoding* in termini di produzione e quindi codifica del parlato emotivo da parte di soggetti appartenenti ad una più lingue, piuttosto che all'analisi acustica vera e propria delle produzioni realizzate al fine di evidenziare e rilevare particolari indici acustici aventi carattere distintivo nelle diverse emozioni trattate. Sono pertanto stati raccolti sotto questa etichetta o categoria solo gli studi che si sono occupati oltre che del *decoding* anche di misurazioni acustiche dirette nel dominio della frequenza, dell'intensità o della durata.

Riassumendo, nel prosieguo verrà pertanto operata una distinzione tra studi volti rispettivamente:

- al *decoding*, ovvero studi in cui viene indagata la capacità degli esseri umani a decodificare il parlato emotivo in una o più lingue diverse dalla propria.
- all'*encoding*, ovvero studi in cui si confrontano due o più lingue alla ricerca dei meccanismi e degli indici acustici interessati nella produzione del parlato emotivo in due o più lingue;
- all'*encoding/decoding*, ovvero studi in cui i due tipi di studio o approccio si integrano e si completano al fine di oggettivare e verificare i risultati derivanti dall'uno e dall'altro.

Per la ricerca di possibili studi condotti a livello cross-linguistico-culturale nel campo delle emozioni vocali secondo gli approcci sopra descritti, oltre a fare riferimento a precedenti studi in cui era già stata effettuata una ricognizione di tali studi²⁰⁶, e per una ricerca degli studi condotti più recentemente, è stata effettuata una ricerca attraverso **MetaBib**, un servizio dell'Ufficio Automazione Biblioteche offerto agli utenti dell'Università della Calabria basato sul software *Metalib*, appartenente alla stessa famiglia di cui fa parte anche *Aleph*, il catalogo online delle biblioteche. Attraverso tale servizio è stato possibile accedere a tutta una serie di *open archives* e risorse *on-line* e consultare i *full-text* delle riviste ricercate.²⁰⁷ Per la ricerca sono state usate parole chiave che potessero in qualche modo richiamare gli studi da ricercare, come ad esempio “*vocal emotion AND cross-linguistic OR cross-cultural*” ecc. Tutte le ricerche sono state effettuate in lingua inglese. Date queste premesse, la rassegna degli studi

²⁰⁶ Tra cui Pavlenko (2005) per gli studi prevalentemente volti al *decoding*, e le meta-analisi condotte rispettivamente da Juslin, Laukka (2003), volta alla ricerca di relazioni tra espressione vocale ed espressione musicale delle emozioni, e da Elfenbein, Ambady (2002), sull'universalità e la specificità culturale del riconoscimento delle emozioni.

²⁰⁷ L'accesso al servizio è possibile solo da postazioni direttamente collegate alla rete di ateneo o tramite il servizio di *ProxyServer*. Per un elenco completo delle risorse a cui si è avuto accesso si consulti il seguente link: <http://www.biblioteche.unical.it/riviste.htm> (data ultima consultazione: giugno 2009).

riportati potrebbe, pertanto, non essere completa o aggiornata per limitazioni dovute ai canali di ricerca utilizzati.

Tenendo a mente le questioni metodologiche a cui si è più sopra dato ampio spazio, si cercherà di mettere in evidenza da un lato quelle che sono state le strategie adottate in ciascuno degli studi qui riportati e, dall'altro le conclusioni e le risultanze che dalle rispettive indagini sono emerse, elementi che costituiranno un necessario punto di partenza per la definizione di quelle che saranno gli obiettivi e le scelte metodologiche operate ed adottate nella presente indagine.

Prima di affrontare nei dettagli gli aspetti a cui si è appena fatto riferimento, verrà di seguito aperta una breve parentesi su quelle che possono essere definite differenze culturali nell'espressione delle emozioni vocali.

7.1 Differenze culturali nell'espressione delle emozioni (vocali)

Come fanno rilevare D'Urso, Trentin (2004), e da quanto già emerso nella breve esposizione sulle teorie delle emozioni (*cf.* § 3.3), si assiste ad una disputa teorica sulla natura delle emozioni che, da un lato, le vede come un fenomeno universale avente le medesime caratteristiche in tutti gli esseri umani indifferentemente dalla cultura di appartenenza, dall'altro, come un fenomeno che trae origine dalla società e dalla cultura in cui queste si manifestano con caratteristiche variabili in ciascun contesto culturale.

Quello che D'Urso, Trentin (2004) chiamano "dilemma natura/cultura" è un problema complesso e molto lontano dal trovare una soluzione chiara ed univoca. Se si esaminano una serie di studi condotti da Scherer e collaboratori negli ultimi 20-30 anni, uno dei pochi (se non l'unico) ad aver presentato indagini trans-culturali in modo esteso e sistematico, si possono scorgere una serie di considerazioni che vanno a favore sia dell'una che dell'altra parte. Si tratta di tre studi volti a valutare il tipo di antecedenti, le reazioni e i meccanismi che precedono e che regolano alcune emozioni.²⁰⁸

Il primo di tali studi è quello di Scherer, Wallbott, Summerfield (1986). Lo studio è stato condotto in otto paesi europei (Francia, Germania, Belgio, Olanda, Italia, Svizzera, Spagna e Israele) su un insieme di 779 soggetti attraverso un questionario a risposte aperte di tipo comparativo su quattro emozioni (rabbia, paura, tristezza e gioia). Da questo studio emergerebbe una notevole omogeneità riguardo gli antecedenti e le reazioni associate a

²⁰⁸ Per non perdere di vista quelli che sono gli obiettivi del presente lavoro, ci si limiterà solo a qualche breve accenno rimandando il lettore alle fonti originali per una visione più dettagliata e all'opera di D'Urso, Trentin (2004) per una discussione più ampia e più critica.

ciascuna delle emozioni esaminate, confermando quindi un certo grado di universalità delle emozioni.

Quello di Scherer, Wallbott, Matsumoto, Kudoh (1988), il secondo, è invece uno studio condotto su due culture (americana e giapponese) con lo stesso questionario utilizzato nel precedente studio e con un sottocampione europeo di confronto ricavato dalla precedente indagine. In questo caso le modalità con cui le emozioni si manifestano sono risultate alquanto simili, supportando l'ipotesi dell'universalità delle emozioni. Tuttavia, sono state riscontrate anche una serie di differenze riconducibili, probabilmente, alla diversa appartenenza culturale dei soggetti (soprattutto occidentale vs. orientale). Si tratterebbe, quindi, di un elemento a favore della specificità culturale delle emozioni, anche se gli autori sottolineano la necessità di un contributo da parte di altri studiosi (antropologi, sociologi ecc.) per spiegare e meglio definire quest'ultimo aspetto.

Il terzo studio chiamato anche ISEAR (*International Study on Emotion Antecedents and Reactions*) e condotto da Wallbott, Scherer (1988), tenta invece di superare alcune limitazioni delle due precedenti indagini utilizzando un set di emozioni più ampio (con l'aggiunta di disgusto, vergogna e colpa) e un questionario questa volta a risposta chiusa basato sulla categorizzazione dei dati scaturita dai due precedenti studi. I dati emersi dallo studio condotto questa volta su 37 nazioni confermano quanto riscontrato nel precedente studio riguardo la presenza di punti in comune e sostanziali differenze nella manifestazione delle emozioni.

In linea con le risultanze dei tre studi richiamati, anche quanto riportato da Mesquita (2007) in un recente articolo, in cui racconta di un'esperienza scientifica avuta qualche anno fa in uno studio-intervista in cui venivano comparate emozioni in contesti Nord-Americani e Giapponesi. L'intervista prevedeva domande quali ad esempio "What was the intensity of your emotion?", domanda che fu però cambiata perché lasciava i soggetti giapponesi attoniti. Da ciò l'autrice comprese:

*"That emotions, as they exist in Japanese contexts, are not located within any single person, but rather emerge within an interaction between people. Since emotions are contextualized and considered to be primarily a property of the relationship, a question that asks for the qualification of an internal state, such as that on emotional intensity, does not fit the Japanese understanding."*²⁰⁹

²⁰⁹ Cfr. Mesquita, B. (2007), Emotions are culturally situated, *Social Science Information*, 46(3): 410.

Risulta evidente, quindi, come quanto più sopra riportato nel § 6.5 riguardo le etichette verbali legate alle emozioni rappresenti una problematica reale e di come vi siano fondamentali differenze nella concettualizzazione delle emozioni nella cultura Giapponese e Occidentale (e non solo). Secondo quanto appena esposto, l'approccio di Scherer si renderebbe perciò più idoneo e più vicino alla cultura occidentale.

La definizione di emozione fornita da Scherer (2005: 697) come:

“an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism”,

conterrebbe, secondo l'autrice, una serie di assunzioni fortemente radicate nella cultura e nei modelli emotivi occidentali. La prima assunzione è da ricondurre al fatto che le emozioni hanno luogo a livello individuale, piuttosto che trascendere l'individualità. La seconda è che l'utilità delle emozioni viene meglio descritta al livello dell'individuo invece che a quello delle relazioni. La terza è che un'emozione viene meglio caratterizzata a livello intra-personale o individuale, piuttosto che essere orientata, ad esempio, agli esiti relazionali. Pur applaudendo e riconoscendo gli sforzi di Scherer, l'autrice suggerisce:

*“that a viable conceptualization of emotion must account for emotional events as they are conceived of and experienced in a wider range of cultural contexts.”*²¹⁰

Le emozioni aiutano gli esseri umani a navigare nelle proprie relazioni sociali e all'interno di queste esistono differenze marcate dove sia gli accordi relazionali, sia i significati che animano e giustificano questi accordi, variano da cultura a cultura. Si tratta di quelli che Mesquita, Leu (2007) hanno definito come *cultural models*, modelli comprendenti accordi e relazioni che non rappresentano solo la normativa o ciò che è prescritto, ma anche e soprattutto abitudini. Anche se tali modelli culturali appaiono invisibili a coloro che li utilizzano e li mettono in atto, questi possono essere resi più espliciti comparando differenti modelli culturali. Riportando numerosi esempi da precedenti studi condotti su confronti tra Nord-americani e Giapponesi, Mesquita (*op. cit.*: 413) suggerisce che una definizione più soddisfacente di emozione dovrebbe incorporare il fatto che *“emotions are contextually constituted”*, aggiungendo inoltre che, piuttosto che essere definite come caratteristiche della

²¹⁰ *Ibidem*, pag. 411.

mente, le emozioni dovrebbero essere collocate all'interfaccia tra mente e contesto, come già riportato nel § 3.2 e che qui richiamato per comodità:

“Emotions are interpersonal events that navigate the social context and are informed by it. All emotional events (perhaps everything more complex than a startle) are intensely meaningful, and meaning is provided by the socio-cultural context.”²¹¹

Nel quadro rappresentato dall'autrice le emozioni non sono disgiunte dalla cultura ma, al contrario, sono da essa costituite.

Un altro esempio di come la cultura giochi un ruolo fondamentale nell'espressione delle emozioni viene fornito da Härtel, Härtel (2005: 685):

“Consider the case where a researcher is looking at the intensity rating of happiness of British and Italian samples on a scale of 1 to 10, where 10 is ‘very intense’. The implied difference could be a result of a real difference in happiness (i.e. the Italian sample is truly happier than the British sample) or due to a culturally based reporting difference (i.e. the Italian sample generally report larger differences in emotional intensity). Knowing that the Italian culture is high on affective expressivity while the British culture is low on this dimension is important in understanding that simply looking at who has the highest mean rating does not tell us which sample is happiest.”

A questo proposito sarà interessante ricordare un altro studio condotto da Babad, Wallbott (1986) riprendendo i risultati di un'altra ricerca (Scherer, Wallbott, Summerfield, 1986) per verificare la veridicità di una serie di stereotipi nazionali. Gli autori hanno potuto di fatto constatare che molti degli stereotipi risultarono confermati, rivelando, in alcuni casi, altrettante sorprese. Nel complesso, ancora una volta, i punti in comune si sono rivelati essere maggiori delle differenze.

È evidente come nella valutazione di produzioni emotive a livello cross-linguistico sia fondamentale tenere conto non solo della cultura dell'*encoder*, ma anche della cultura del *decoder*.

Infine, per quel che attiene alle emozioni vocali, vale la pena di ricordare come in questo quadro assuma notevole rilevanza quanto formulato da Scherer (1989) riguardo alle caratteristiche più salienti delle espressioni vocali (ma anche delle espressioni facciali): entrambe risultano determinate sia da “*push effects*”, la mera esternalizzazione di processi

²¹¹ *Ibidem*, pag. 413.

naturali interiori dell'organismo determinati a livello biologico, e da “*pull effects*”, un insieme di norme e modelli determinati e acquisiti a livello socioculturale che definiscono le caratteristiche dei segnali richiesti dai codici socialmente condivisi e utilizzati per la comunicazione di stati interiori nonché di intenzioni comportamentali.

Anche in questo caso Scherer (1994: 189) sottolinea e ribadisce come:

“A comprehensive cross-culturally approach is imperative if we want to even begin to understand the complex interrelationship between push and pull factors in the production of vocal/ facial affect expressions.”

Nel rispetto dei meccanismi di “*push*” e “*pull effects*”, Scherer assume, tuttavia, una predominanza di “*push effects*” nella fase iniziale della comparsa dell'evento emotigeno successivamente sopraffatti con estrema rapidità da “*pull effects*”, soprattutto in determinati contesti sociali.

7.2 Decoding delle emozioni a livello cross-linguistico

Gli esseri umani sono in grado di identificare correttamente specifiche emozioni codificate in una frase o in un messaggio verbale anche in una lingua ad essi sconosciuta.

Si tratta di un dato di fatto emerso da una serie di studi che sono stati condotti negli anni passati e per i quali si registra il maggior numero di ricerche condotte.

Andando a ritroso nel tempo, il primo studio condotto in questa direzione²¹², confermando e offrendo un ottimo spunto di partenza ai successivi studi che si sono poi succeduti negli anni, è quello di Kramer (1964), che chiese a 27 parlanti americani di madrelingua inglese di identificare produzioni verbali di cinque emozioni (*anger, love, contempt, indifference, grief*). Le produzioni emotive di uno stesso set di frasi, realizzate da attori americani e giapponesi, sono state somministrate agli ascoltatori in ciascuna delle due lingue, con le produzioni per l'inglese presentate anche in una versione filtrata (*content filtered*)²¹³. Dalle percentuali di corretta identificazione rilevate, Kramer concluse che gli ascoltatori erano in grado di identificare le produzioni emotive con un'accuratezza al di sopra della pura casualità con diversi gradi di riconoscimento per le varie emozioni, anche in quei casi in cui la lingua, in questo caso il giapponese, non era nota agli ascoltatori: si arrivava in

²¹² Vale a dire l'utilizzo e la somministrazione di stimoli in una lingua diversa da quella dei soggetti ascoltatori.

²¹³ In tutti quei casi in cui vengono utilizzate produzioni *content-filtered* lo scopo è quello di rimuovere qualsiasi informazione di tipo linguistico lasciando inalterati quelli che sono i contorni prosodici delle produzioni somministrate agli ascoltatori.

media al 70% per la stessa lingua degli ascoltatori (inglese) nella versione non filtrata e al 61% per la versione filtrata, mentre per le produzioni nella lingua non nota agli ascoltatori (giapponese) la percentuale media di corretto riconoscimento arrivava al 58%.

In un altro studio Beier, Zautra (1972) chiesero a studenti americani, polacchi e giapponesi di giudicare campioni di parlato emotivo (*angry, sad, happy, flirtatious, fearful, indifferent*) di lunghezza differente (*hello, good morning, how are you*, e una frase intera) prodotte in inglese da un soggetto americano. I risultati dimostrarono, in questo caso, la presenza nel canale vocale di indici emotivi universali e che le percentuali di corretto riconoscimento a livello interculturale aumentavano con l'aumentare della lunghezza dei campioni di parlato presentati. Gli ascoltatori americani nativi rivelarono, rispetto agli altri due gruppi, una maggiore capacità nel riconoscere emozioni vocali in stimoli più brevi (*in-group advantage*); a livello di riconoscimento cross-culturale, invece, le percentuali di corretta identificazione si riducevano tra i gruppi all'aumentare della lunghezza (in termini di durata) degli stimoli, fino ad annullarsi quasi completamente nel caso dello stimolo emotivo rappresentato dalla frase intera.

McCluskey, Albas (1981) presentarono stimoli vocali filtrati, prodotti rispettivamente in canadese (inglese) e in messicano (spagnolo) da tre soggetti femminili²¹⁴, a due gruppi di ascoltatori (anche questi uno canadese e uno messicano). Ne scaturì che gli ascoltatori messicani erano più abili a riconoscere le emozioni negli stimoli presentati di quanto non lo fossero i soggetti dell'altro gruppo (canadese), anche quando gli stimoli provenivano da parlatori canadesi. La maggiore abilità da parte dei soggetti messicani rispetto ai canadesi emersa dall'esperimento condotto da McCluskey, Albas (1981), viene spiegata da Dromey, Silveira, Sandor (2005: 356) alla luce del fatto che il cervello valuta i contorni prosodici a livello di frase in modo diverso da quanto accade nella lingua parlata e che i meccanismi di interpretazione della prosodia nel parlato naturale possono differire da quelli messi in atto nel caso di stimoli filtrati, sottolineando come:

[...] extensive modulation of natural speech to the point of masking linguistic information may not allow the language centers of the brain to be operating to the extent that we intend to test them."

L'abilità dell'uomo ad identificare stimoli vocali emotivi in una lingua diversa dalla propria è stata accertata anche in soggetti non adulti con una serie di studi condotti da

²¹⁴ Le emozioni considerate erano in questo caso *anger, happiness, love, sadness*. Ai parlatori era stato chiesto di pronunciare due frasi che venivano loro in mente per illustrare ciascuno degli stati emotivi richiesti.

McCluskey e suoi collaboratori, come ad esempio in McCluskey *et al.* (1975), in cui erano stati utilizzati bambini piuttosto che soggetti adulti, rilevando, per l'appunto, tassi di riconoscimento cross-culturale in alcuni casi anche superiori al riconoscimento intra-culturale. Sempre con la stessa tecnica, ma utilizzando 6 parlatori maschili per la produzione di stimoli in inglese ed altrettanti per le produzioni in creolo, Albas, McCluskey, Albas (1976), rilevando una performance superiore da parte degli ascoltatori a livello di riconoscimento di emozioni prodotte da membri appartenenti al proprio gruppo, suggerirono che lingua e cultura erano da considerarsi fattori cruciali nella trasmissione delle emozioni anche a livello non-verbale, ammettendo, tuttavia, la difficoltà di interpretazione dei risultati per l'aver utilizzato, nella produzione degli stimoli emotivi, materiale non controllato.²¹⁵ Scherer *et al.* (2001:77) fanno qui giustamente osservare quanto segue:

“The speakers may have used culture-specific expressions, possibly with characteristic suprasegmental cues (intonation contours, rhythm), that might be difficult for outsiders to decode.”

Sempre su non adulti si sono concentrati Roseberry-McKibbin, Brice (1999)²¹⁶, che chiesero a 9 bambini messicani (di età compresa tra 5 e 7 anni) di giudicare emozioni espresse in L1 spagnola e L2 inglese, in cui avevano una competenza limitata, constatando che i bambini erano nettamente migliori in L1 spagnola che in L2 inglese. Tuttavia, Pavlenko (2005: 63) fa notare:

“Unfortunately, it is unclear whether these results are attributable to differences in proficiency or to the fact that the male speaker of English was not as effective as the female speaker, whose expressions the children decoded as well those of Spanish speakers.”

L'esperimento condotto da Solomon, Ali (1975) con un gruppo parlanti inglese nativi (americani) e con un gruppo di parlanti inglese come L2 (indiani), tutti studenti compresi tra i 14 e i 18 anni, mette in evidenza come il grado di attenzione e la sensibilità degli ascoltatori nel riconoscere correttamente espressioni emotive sulla base del contenuto verbale o, diversamente, sulla base degli indici acustici veicolati attraverso il canale vocale, varia non solo al variare della comunità linguistica di appartenenza, ma anche al variare della

²¹⁵ Anche in questo caso, come in McCluskey, Albas (1981), ai soggetti era stato di esprimere *anger, happiness, love, sadness* utilizzando qualsiasi parola veniva loro in mente.

²¹⁶ *Op. cit.* in Pavlenko, A. (2005), *Emotions and Multilingualism*, New York: Cambridge University Press, pag. 63.

competenza che si ha della lingua presentata (L1 vs. L2). Nel caso di parlanti L2 (studenti indiani), secondo Solomon, Ali, nell'interpretazione verbale di stati emotivi²¹⁷ vi sarebbe un maggiore ricorso a quello che è rappresentato dal contenuto verbale di un enunciato, soprattutto agli stadi iniziali dell'apprendimento della lingua straniera, mentre nel caso di parlanti L1 (studenti americani), dall'analisi dei risultati, si è evidenziato, da parte di questo gruppo, un ricorso predominante agli indici acustici dell'emozione veicolati nel canale vocale, soprattutto all'intonazione degli enunciati.

Un altro studio che viene citato in tutti i lavori sulla decodifica di parlato emotivo è quello di Van Bezooijen *et al.* (1983) in cui è stata studiata la capacità di ascoltatori provenienti dall'Olanda, da Taiwan e dal Giappone a riconoscere emozioni vocali prodotte in lingua olandese.²¹⁸ La capacità degli ascoltatori a riconoscere le produzioni presentate ha evidenziato un grado di accuratezza nettamente al di sopra del caso per gli ascoltatori olandesi, ma anche per i due gruppi asiatici che non avevano mai avuto contatti con la lingua olandese.

In Magno Caldognetto, Kori (1983), invece, sono stati utilizzati due set di tre vocali, /a/, /e/, /o/, prodotte rispettivamente da un parlante italiano e da una parlante giapponese²¹⁹ con sei intonazioni emotive (*joy, fear, surprise, sadness, disgust, anger*). Le risultanti produzioni sono successivamente state somministrate a due gruppi di 24 ascoltatori (uno italiano e uno giapponese equamente suddivisi tra maschi e femmine e senza alcuna conoscenza dell'altra lingua presentata) in tre diversi *tests*:

1. uno di identificazione (di tipo *forced-choice*);
2. uno di giudizio di adeguatezza;
3. uno di "etichettatura" libera delle espressioni emotive (proposto solo agli ascoltatori giapponesi per le espressioni in italiano).

Dall'analisi dei dati ricavati dai *tests* somministrati agli ascoltatori Magno Caldognetto, Kori rilevarono che l'espressione di alcune emozioni (*joy, surprise, disgust*) era specifica di ciascuna delle due culture rendendone così difficile il riconoscimento da parte di membri appartenenti ad una cultura diversa da quella del parlante. Allo stesso tempo, però, rilevarono la presenza di espressioni riconoscibili in entrambe le culture, confermando come

²¹⁷ Nel presente esperimento sono state utilizzate frasi e commenti simili a quelli utilizzati dall'insegnante codificate in tre categorie emotive (*positive, neutral e negative*) e tre intonazioni (*pleased, indifferent e displeased*).

²¹⁸ In questo caso la frase presentata era "two months pregnant" nelle emozioni di *disgust, surprise, shame, joy, fear, contempt, sadness, anger* con l'aggiunta di una produzione non emotiva, prodotte da 4 soggetti di sesso maschile e 4 di sesso femminile.

²¹⁹ I due parlanti erano, rispettivamente, un attore professionista nel caso dell'italiano e un'attrice di teatro non professionista nel caso dello giapponese.

vi siano elementi a supporto delle somiglianze a livello cross-culturale, ma anche elementi che ne supportano la specificità culturale. Nel caso di stimoli presentati nella stessa lingua degli ascoltatori le percentuali medie di corretta identificazione riscontrate vanno dall'87% per gli italiani al 77% per i giapponesi. Nel caso di stimoli presentati in lingua diversa da quella degli ascoltatori è stata rilevata, invece, una minore abilità nell'identificare correttamente gli stimoli da parte degli ascoltatori italiani (53%) rispetto ai giapponesi (81%).

Sostanzialmente diverso è stato l'approccio utilizzato in Rintell (1984) che presentò a due gruppi di ascoltatori, uno di madrelingua inglese e uno di madrelingua diversa dall'inglese²²⁰, undici conversazioni rappresentanti sei emozioni²²¹. La caratteristica principale di dette conversazioni stava nel fatto che il parlante, nel raccontare ad un altro soggetto di un'esperienza che gli aveva fatto vivere una particolare emozione, raccontava della storia vissuta senza mai fare riferimento o menzione dell'emozione esperita o provata. Ciò che emerse fu che la scelta dell'emozione corretta, basata sul mero ascolto delle conversazioni, risultò alquanto difficile in entrambi i gruppi: i soggetti di madrelingua diversa dall'inglese non traevano alcun vantaggio dal contenuto verbale dei testi, mentre i parlanti nativi di madrelingua inglese erano, in questo caso, significativamente più abili (76%) rispetto ai parlanti non nativi (33%), con difficoltà maggiori nello svolgimento del *task* riscontrate da parte del gruppo cinese. Una possibile spiegazione alle difficoltà riscontrate da parte dei due gruppi è da ricercare, secondo la ricercatrice, nel fatto che le espressioni emotive sono da considerarsi uno degli atti illocutori veri e propri, in quanto l'espressione di una emozione rappresenta un atto comunicativo in cui l'ascoltatore riconosce l'intenzione del parlante a compiere quel determinato atto. Così intese, le produzioni verbali che esprimono determinate emozioni realizzano simultaneamente una serie di funzioni sociali e psicologiche. Poiché l'espressione delle emozioni è anche una delle funzioni pragmatiche, la Rintell sottolinea come determinate norme regolino la possibilità di esprimere determinate emozioni in modo diretto il cui livello di intensità comunicativa viene modulato attraverso l'uso del linguaggio: attraverso il linguaggio il parlante è in grado di minimizzare o massimizzare il grado di intensità di una specifica emozione che intende trasmettere durante un atto comunicativo.

Graham *et al.* (2001) investigarono, invece, l'abilità di parlanti nativi inglesi e di parlanti nativi spagnoli e giapponesi (apprendenti la lingua inglese come L2) nel riconoscere

²²⁰ Nel gruppo di madrelingua diversa dall'inglese, in cui l'inglese veniva, invece, appreso come L2, erano presenti 66 parlanti spagnoli, 20 arabi e 17 cinesi.

²²¹ Il range di emozioni era dato da *pleasure, anger, depression, anxiety, guilt* e *disgust*.

otto espressioni emotive in lingua inglese prodotte da attori professionisti.²²² Anche qui i risultati confermarono sostanziali differenze nell'abilità da parte di ascoltatori nativi e non nativi ad identificare emozioni espresse attraverso la voce. Furono, infatti, riscontrate percentuali di corretta identificazione maggiori nel gruppo di parlanti nativi (in media del 59%) rispetto a quelle del gruppo dei non nativi composto da spagnoli e giapponesi (rispettivamente del 42% e del 39%). In tutti e tre i gruppi l'emozione maggiormente identificata è stata quella di *anger*. Dal confronto delle matrici di confusione gli autori hanno anche rilevato differenze sistematiche tra spagnoli e giapponesi rispetto ai nativi inglesi.²²³ Inoltre, l'analisi dei risultati conseguiti da parte di ascoltatori non nativi a diversi livelli di competenza della lingua inglese non hanno evidenziato significativi miglioramenti nell'abilità di giudicare e riconoscere espressioni emotive in lingua inglese all'aumentare del livello di competenza, escludendo perciò qualsiasi tipo di correlazione nell'abilità a riconoscere emozioni con il livello di competenza di L2. Ciò dimostrerebbe, secondo gli autori, come l'abilità nell'identificare correttamente emozioni trasmesse attraverso la voce in una lingua seconda rappresenti un processo molto lento e che non può essere acquisita da parlanti di L2 se non attraverso una lunga esposizione a quelli che sono i contesti nativi della lingua appresa o attraverso una particolare attenzione in un contesto didattico ed educativo volta allo sviluppo di tali abilità. Quanto detto giustificherebbe anche le difficoltà riscontrate dai due gruppi di ascoltatori in Rintell (1984), nonché l'assoluta correttezza di quanto da lei suggerito sulla funzione delle emozioni e sulla presenza di determinate norme a cui, a livello didattico, non sempre viene data la dovuta attenzione come suggeriscono appunto Graham *et al.* (2001).

Una serie di altri studi condotti con soggetti americani e giapponesi confermano che il riconoscimento del parlato emotivo risulta affetto da differenze culturali sia da parte del parlante sia da parte degli ascoltatori. Ad esempio, Erickson, Maekawa (2001), con stimoli in lingua inglese somministrati ad ascoltatori americani e giapponesi, riportano percentuali di corretta identificazione che oscillano tra il 90 e il 100% per la stessa lingua e tra il 64 e il 92% per la lingua straniera. Valori più o meno simili vengono riportati da Nakamichi *et al.* (2002) con l'utilizzo di frasi non filtrate tratte da due film (sia in giapponese che in inglese) e due gruppi di ascoltatori (anche qui giapponese e inglese). Shigeno (1998, 2003)²²⁴ ha messo in evidenza le somiglianze e le differenze nel riconoscimento delle emozioni da parte di soggetti

²²² Diversamente da altri studi sino a qui riportati, in tal caso si è fatto utilizzo di monologhi sotto forma di narrazione.

²²³ Questa sistematicità suggerirebbe di fatto, secondo gli autori, una maggiore differenza di indici prosodici e paralinguistici nell'espressione vocale tra lingue tipologicamente distinte.

²²⁴ Sebbene si tratti di due studi che hanno investigato le caratteristiche bimodali uditivo-visive della percezione delle emozioni risultano qui interessanti i risultati relativi al solo canale uditivo.

Giapponesi e Nord-Americani utilizzando sia espressioni vocali che facciali prodotte da attori Americani e Giapponesi, allo scopo di trasmettere sei emozioni base (gioia, sorpresa, rabbia, disgusto, paura, tristezza). All'esperimento hanno preso parte due gruppi di soggetti, sette giapponesi nativi e cinque nord-americani nativi ai quali sono stati presentati, dapprima solo gli stimoli uditivi, poi solo gli stimoli visivi ed infine gli stimoli audio-visivi frutto del montaggio. I risultati conseguiti hanno dimostrato che: a) i soggetti riconoscono meglio espressioni vocali appartenenti ad un soggetto della propria cultura; b) i soggetti giapponesi non sono performanti nel riconoscere la "paura" negli stimoli uni-modalità visivi e uditivi, e c) entrambi i soggetti americani e giapponesi riconoscono stimoli bi-modalità audio-visivi incongruenti molto più spesso sulla base dello stimolo visivo che non dello stimolo uditivo. Secondo l'autore i risultati dimostrano quanto sia difficile riconoscere le emozioni di un parlante appartenente ad una cultura diversa e che gli individui usano prevalentemente il canale visivo per identificare le emozioni.²²⁵ Ancora una volta risulta evidente come il grado di identificazione per alcune emozioni risulti maggiore per parlanti e ascoltatori aventi lo stesso background culturale.

Più o meno simile, per quanto riguarda lo studio delle emozioni vocali a livello cross-linguistico con l'ausilio di espressioni facciali, è l'esperimento condotto da Tickle (1999, 2000) in cui l'autrice si chiede fino a che punto le emozioni siano dovute a meccanismi di risposta psico-biologica (*psycho-biological response mechanisms, PBRMs*), e quindi quasi universali, e fino a che punto, invece, siano dovute a convenzioni sociali. Per rispondere a questi quesiti approccia il problema da un punto di vista cross-linguistico affrontando dapprima una serie di complicazioni metodologiche, descrivendo e motivando le procedure adottate per ovviare a tali difficoltà. Nello studio intrapreso Tickle usa espressioni *non-sense* ed espressioni facciali di emozioni quasi-universalmente riconosciute al fine di risolvere *a-priori* problemi di traduzione, problemi etici nella raccolta dei dati, con un giusto compromesso tra artificialità del dato e coerenza, mascherando il contenuto verbale delle espressioni e consentendo comunque di mettere in evidenza eventuali influenze esercitate da specifiche qualità legate a singole vocali. Prendendo in esame soggetti femminili inglesi e giapponesi, in quanto lingue non correlate tra di loro, l'autrice ha chiesto ai soggetti di simulare determinate emozioni (*happy, sad, angry, fearful e calm*) secondo una particolare

²²⁵ Quest'ultimo dato rispecchia in un certo qual modo la *3Vs equation* proposta da Mehrabian (1981) a cui si è fatto più sopra riferimento (vedi § 4).

procedura.²²⁶ Il materiale raccolto è stato validato da due gruppi di ascoltatori, uno per lingua, al fine di selezionare le migliori produzioni che sono poi state somministrate con un test di tipo *forced-choice* con cinque opzioni a due gruppi di ascoltatori appartenenti alle due lingue studiate. Andando a quelli che sono i risultati dell'esperimento condotto dalla Tickle, sebbene anche in questo caso venga messa in luce la necessità di ulteriori approfondimenti, è sostanzialmente emerso come soggetti inglesi e soggetti giapponesi siano stati in grado di decodificare produzioni vocali emotive nella propria lingua (60% di accuratezza per i primi, 36% per i secondi) e che soggetti inglesi e soggetti giapponesi sono stati in grado di decodificare produzioni emotive in lingua diversa dalla propria, con percentuali di accuratezza rispettivamente del 35% e del 42%. Da quanto evidenziato dalla Tickle nella ricerca condotta, emergerebbe come vi sia un'influenza culturale con specifiche norme di esibizione sociale maggiore nel caso del giapponese, molto di più di quanto non lo sia per l'inglese nell'espressione di specifiche emozioni.

Restando nell'ambito degli esperimenti condotti con soggetti americani ed asiatici Sawamura *et al.* (2007) hanno intrapreso quello che gli stessi autori hanno definito come *cognition experiment* utilizzando soggetti giapponesi, americani e cinesi senza esperienze di vita all'estero. Per l'esperimento è stato utilizzato un database di parlato emotivo senza informazione linguistica²²⁷ le cui produzioni sono state somministrate agli ascoltatori in due diversi esperimenti: nel primo è stato chiesto agli ascoltatori di valutare liberamente le produzioni presentate (*anger, joy e sad*) sulla base di tre componenti emotive (*anger, joy e sad*) attribuendo a ciascuna valutazione relativa alle componenti emotive un punteggio (*score*) da 5 (*strongly perceived*) a 1 (*nothing perceived*) sul grado di percezione dell'emozione presentata; nel secondo esperimento, utilizzando questa volta solo soggetti cinesi, è stato chiesto agli ascoltatori di valutare le stesse produzioni sulla base di sei componenti emotive (*anger, joy, sadness, fear, surprise e disgust*) e di esprimere un grado di percezione dell'emozione presentata seguendo la precedente scala di giudizio. Senza scendere in ulteriori dettagli, dalle analisi condotte è emerso che la maggior parte delle produzioni veniva percepita come avente componenti emotive multiple, anche quando i soggetti esprimevano una singola emozione, e che la PCA (*principle component analysis*) da loro condotta implica

²²⁶ Per i dettagli sul protocollo utilizzato per la fase di *encoding* del materiale utilizzato nel presente esperimento si rimanda a Tickle (1999), pag. 307.

²²⁷ Il database in questione è quello di Sakuraba *et al.* (2000) contenente la parola "Pikachu" prodotta da bambini da 4 a 10 anni con 4 emozioni diverse (*anger, joy, sad, e surprise*) nel vedere una scena emotiva che coinvolgeva il protagonista "Pikachu" del cartone animato "Pocket Monster".

che gli esseri umani siano in grado di percepire le emozioni a partire da stimoli vocali senza informazione linguistica con un grado di accuratezza del 60% circa.

Shochi, Aubergé, Rilliard (2007) riportano, invece, i risultati di un esperimento percettivo sul riconoscimento da parte di giapponesi, francesi e inglesi (americani) di espressioni emotive in lingua giapponese. Agli ascoltatori²²⁸ è stato chiesto di scegliere l'attitudine espressa da un soggetto giapponese per 7 frasi (da 1, 2, 5 o 8 more) prodotte con 12 diverse attitudini.²²⁹ Nella somministrazione dei test percettivi ai tre gruppi di ascoltatori²³⁰ (introdotti da una lunga spiegazione nella lingua del gruppo con numerosi esempi) i risultati del gruppo giapponese sono stati utilizzati come validazione del corpus. Dai risultati è emerso che i soggetti giapponesi riconoscono tutte le produzioni con percentuali di corretta identificazione superiori al caso seppur con qualche confusione tra alcune espressioni (in questo caso due espressioni di cortesia). I francesi e gli americani non riconoscono, invece, le due espressioni giapponesi di cortesia, soprattutto in un caso (*kyoshuku*, in quanto non esistente nelle espressioni di cortesia nella società occidentale) che è stata sistematicamente confusa con arroganza o irritazione.

Poiché la struttura ritmica delle lingue può variare notevolmente, per Dromey, Silveira, Sandor (2005) è assai ragionevole che soggetti cresciuti in presenza di ambienti con più lingue, sembrano essere o sembrano aver sviluppato una particolare capacità nell'interpretazione dei diversi indici acustici, non solo riguardo a ciò che ascoltano, ma anche riguardo ai *patterns* prosodici della propria lingua. L'idea dello studio intrapreso dagli autori nasce dall'osservazione che alcuni soggetti all'interno del gruppo di ascoltatori etnicamente diversi,

“were more accurate than others during a listening task where they were asked to choose which emotion was being conveyed by an English speaker.”²³¹

²²⁸ Sia i francesi che gli inglesi non avevano né conoscenze né competenze nella lingua giapponese.

²²⁹ Cfr. Shochi, Aubergé, Rilliard (2007), pag. 2097. Gli autori usano qui il termine attitudine al posto di emozione in quanto: *“Attitudes (as “labels” or “notions”) are sometimes confused with Emotions (as “labels” or “notions”) since both use specialized prosodic encoding in the acoustic signal. Since the prosodic emotional expressions are carried in parallel with the prosodic linguistic structures, then we can view expressions of attitudes as an integral part of “language interaction building” (i.e. communication). Some affects such as surprise, are classified either as attitudes or emotions. Our position is to assume that surprise can be an emotion if it is involuntarily processed, and can be an attitude when it is voluntarily processed and needs to be learned. [...] Some attitudes can be expected to be universal in part, both in terms of what is expressed and how it is expressed. But because attitudes are socially constructed for and by the language, they can exist in one language but not another; moreover the prosodic realization of one specific attitude in a specific language may not be recognized (or may be ambiguous) in another language.”*

²³⁰ Ai quali era concesso un solo ascolto per stimolo.

²³¹ Cfr. Dromey, Silveira, Sandor (2005), pag. 353.

decidendo così di aumentare il gruppo di ascoltatori allo scopo di valutare,

*“the effect of language background on the performance of the task. The goal was to measure accuracy rates in the perception of affective prosody across listeners with diverse linguistic backgrounds compared with English mother tongue listeners.”*²³²

È stata, quindi, studiata la capacità di soggetti di *English Mother Tongue* (EMT), ovvero di madrelingua inglese, differenziati in poliglotti (EMT_p) e monoglotti (EMT_m)²³³, e di *Other Mother Tongue* (OMT), ovvero di altra madrelingua, a riconoscere parole prodotte con una intonazione neutra e di collera (*anger*). Le parole (da 1 a 3 sillabe) sono state registrate da un attore professionista EMT_m. Durante la fase di *testing* le opzioni fornite all'ascoltatore erano le 5 emozioni più l'opzione neutro, nonostante ci fossero in realtà solo due stati emotivi di cui appunto uno neutro e l'altro collera. Gli ascoltatori aventi EMT_p hanno conseguito risultati più alti rispetto agli ascoltatori OMT, ma non gli EMT_m. Nei tre gruppi di ascoltatori, inoltre, le donne sono risultate più accurate nei riconoscimenti rispetto agli uomini. È probabile, secondo gli autori, che l'apprendimento di una seconda lingua abbia in qualche modo aiutato gli EMT_p a sviluppare capacità percettive aggiuntive nel decodificare emozioni nella propria lingua nativa.

Uno degli studi ritenuti tra i più ampi a livello di *decoding* cross-culturale delle emozioni vocali è, invece, lo studio condotto da Scherer, Banse, Wallbott (2001) in cui vengono riportati, analizzati e confrontati i risultati dell'accuratezza di riconoscimento di espressioni vocali emotive in tre continenti (Europa, Nord America e Asia). Il materiale presentato agli ascoltatori è stato prodotto da quattro attori radiofonici professionisti (due uomini e due donne) nelle cinque emozioni *basic*²³⁴ attraverso procedure di elicitazione pilotate. Gli stimoli, due frasi contenenti elementi fonemici di varie lingue europee, sono stati composti da un fonetista esperto: *“Hat sundig pron you venzy.”* e *“Fee gott laish jonkill gosterr.”*²³⁵ Scopo del lavoro è stato quello di verificare l'ipotesi secondo cui le emozioni espresse attraverso la voce (codificate da un soggetto di una particolare cultura) possono essere riconosciute con un certo grado di affidabilità da soggetti di una cultura diversa.²³⁶ Sulla base di precedenti studi sull'espressione delle emozioni a livello facciale e vocale, e

²³² *Ibidem.*

²³³ L'aggiunta delle lettere minuscole in pedice è una mia semplificazione a fini espositivi.

²³⁴ Ovvero le cinque emozioni definite come basic da Ekman (1982): *happiness, sadness, fear, anger e disgust*. Sono state anche considerate ed aggiunte due produzioni neutre per ciascun attore.

²³⁵ Si tratta, in questo caso, di pseudo-frasi composte da sette sillabe, ma con la caratteristica di essere prive di significato semantico.

²³⁶ Il termine “cultura” è qui inteso nell'accezione di paese.

sulla base di assunti teorici riguardo la continuità filogenetica dei segnali emotivi, gli autori hanno ipotizzato che ascoltatori, indipendentemente dalla cultura e dalla lingua di appartenenza, siano in grado di identificare emozioni prodotte vocalmente con risultati al di sopra delle aspettative. A livello cross-linguistico il presente studio, con qualche altro studio tra quelli qui riportati, è uno dei pochi ad aver effettuato un test di validazione in cui è stato chiesto ad un gruppo di 29 partecipanti (tutti studenti tedeschi) di valutare gli stimoli in base a ciascuna delle emozioni su una scala di intensità da 0 (per nulla) a 6 (intensa) e su una scala di naturalezza da 0 a 6.²³⁷ Durante la fase di decodifica agli ascoltatori è invece stato chiesto di focalizzarsi sull'emozione, piuttosto che sul significato della frase. Ai soggetti ascoltatori è stata anche data l'informazione secondo cui le espressioni potevano esprimere più di un'emozione, dando così facoltà agli stessi di scegliere due emozioni per un dato stimolo se credevano che si trattasse di un'emozione mista.²³⁸ Il risultato a cui gli autori sono giunti riporta che gli ascoltatori di 9 paesi (Germania, Svizzera, Gran Bretagna, Olanda, Francia, Italia, Spagna, Stati Uniti, Indonesia) parlanti differenti lingue, sono in grado di identificare le 4 emozioni e gli stati neutri presentati in espressioni vocali di frasi prive di contenuto con un grado di accuratezza superiore al caso, con risposte tra le emozioni ed errori nelle matrici di confusione simili in tutti i paesi. Dai risultati riportati in Tabella 13,

	<i>Neutral</i>	<i>Anger</i>	<i>Fear</i>	<i>Joy</i>	<i>Sadness</i>	<i>Total</i>	<i>Cohen's Kappa</i>
Germany	88	79	74	48	80	74	.67
Switzerland	71	79	70	55	71	69	.62
Great Britain	67	83	70	40	82	68	.62
Netherlands	77	86	65	45	69	68	.60
United States	66	80	72	46	73	68	.60
Italy	81	72	77	39	68	67	.57
France	70	69	71	51	67	66	.57
Spain	69	73	65	30	71	62	.52
Indonesia	70	64	38	28	58	52	.39
Mean across Western countries except Germany	72	77	70	44	72	67	
Mean across all countries ^a	73	76	66	42	71	66	

Tabella 13: Percentuali di riconoscimento degli stimoli emotivi vocali nello studio condotto da Scherer, Banse, Wallbott (2001)²³⁹

²³⁷ La presente affermazione deriva dal fatto che la procedura di validazione degli stimoli presentati agli ascoltatori risulta esplicitata o menzionata solo in pochissimi studi. Ne consegue che, ove non diversamente segnalato nella presente esposizione, o la procedura di validazione non ha proprio avuto luogo, oppure per questioni di ristrettezza e di spazio a disposizione nelle varie riviste si è preferito non darne menzione nei rispettivi resoconti di ricerca. La procedura di validazione risulta di rilevante importanza nel caso di studi di tipo cross-linguistico in quanto consente di selezionare solo le produzioni più rappresentative. Non sarebbe d'altronde ipotizzabile avere identificazioni corrette da parte di ascoltatori stranieri se già gli stimoli non vengono correttamente identificati da soggetti nella cui lingua gli stimoli sono stati codificati. Nello studio di Scherer, Banse, Wallbott (2001), infatti, per effetto della validazione degli stimoli, il "disgusto" è stato rigettato perché le valutazioni medie erano troppo basse e perché vi erano eccessive confusioni con le emozioni di "rabbia" e di "tristezza".

²³⁸ La risposta era obbligatoria anche in caso di incertezza.

²³⁹ Tabella adattata da Scherer, Banse, Wallbott (2001), pag. 85, table 4.

emerge come le lingue con i più alti gradi di accuratezza sono state quelle di origine germanica (olandese e inglese) seguiti da quelle di origine romanza (italiano, francese e spagnolo). I valori più bassi sono, invece, stati rilevati per l'unica lingua non appartenente al ceppo indo-europeo, ovvero l'indonesiano. Se nello studio appena esposto risulta chiaro che la differenza di lingua tra *encoder* e *decoder* giochi un ruolo fondamentale, non è affatto chiara la natura dei meccanismi coinvolti che portano gli autori ad affermare che:

*"It seems that as soon as vocal expressions, other than pure nonlinguistic affect bursts, are used, segmental and suprasegmental aspects of language affect encoding and decoding of emotion."*²⁴⁰

Gli autori attribuiscono queste influenze all'aver utilizzato frasi prive di contenuto composte da unità fonologiche provenienti da diverse lingue indoeuropee, e che tali effetti sono da ricercare o a informazioni segmentali (frequenza fondamentale, differenze di articolazione, strutture formantiche o simili) o a parametri sopra-segmentali come gli indici prosodici (intonazione, ritmo e *timing*). Gli autori non escludono nemmeno la possibilità che, oltre alla dissimilarità linguistica, vi possano essere effetti da ricondurre alla cultura occidentale. Ad ogni modo, quali che siano questi fattori, in tutti i gruppi di ascoltatori è stata riscontrata una percentuale di riconoscimento delle produzioni emotive presentate decisamente assai superiore al semplice caso, con una media complessiva tra tutti gli ascoltatori del 66%.

Secondo un recentissimo studio condotto da Pell *et al.* (2009) l'abilità a comprendere e riconoscere emozioni nel parlato sarebbe da ritenersi indipendente dall'abilità linguistica dei soggetti, coinvolgendo così quelli possono essere definiti *principi universali*, ma sarebbe, allo stesso tempo, anche determinata da variabili linguistiche e culturali. Confrontando come parlanti monolingui argentini (spagnolo)²⁴¹ riconoscono emozioni *basic* a partire da pseudo-frasi (*nonsense speech*) prodotte nella loro lingua da un parlante nativo e in tre lingue straniere (inglese, tedesco e arabo) da parlanti nativi²⁴², è infatti emerso, confermando i risultati di precedenti studi che più sopra sono stati richiamati, come soggetti spagnoli siano stati in grado di riconoscere emozioni espresse nelle diverse lingue con un accuratezza superiore al caso dimostrando un *in-group advantage* nel decodificare espressioni emotive prodotte da parlanti aventi la loro stessa lingua: complessivamente il 64% di corretto riconoscimento di emozioni nella stessa lingua (spagnolo), 59% per l'arabo, 58% per l'inglese

²⁴⁰ Cfr. Scherer, Banse, Wallbott (2001), pag. 88.

²⁴¹ Con poca o indiretta esperienza con parlanti inglesi, tedeschi e arabi.

²⁴² Tutte le produzioni presentate agli ascoltatori spagnoli sono dapprima state validate con gruppi di parlanti nativi di ciascuna lingua in esame. Per maggiori dettagli si rimanda a Pell *et al.* (2009), pag. 111 e segg.

e il 56% per il tedesco.²⁴³ I risultati di questo studio sono inoltre in contrasto con quanto affermato da Scherer, Banse, Wallbott (2001) sul fatto che la capacità di discernere e riconoscere parlato emotivo sia legata a fattori come quelli della diversa tipologia linguistica, con riconoscimenti corretti più elevati in lingue tipologicamente più prossime a quella degli ascoltatori. Da quanto osservato da Pell *et al.* (2009), sebbene gli ascoltatori spagnoli (92%) ritenessero gli stimoli emotivi in lingua araba più difficili da categorizzare, il dato non ha avuto riscontro nella loro reale performance.

Di seguito viene riportata una tabella riassuntiva degli studi citati in questo paragrafo. Rispetto alla versione originale proposta da Pavlenko (2005), la tabella è stata integrata ed aggiornata con una serie di altri studi e ricerche che qui sono stati, per quanto possibile, illustrati e commentati. Il criterio di inclusione adottato è quello specificato più sopra nel tentativo di ripartizione delle ricerche in ambito di *decoding* di emozioni vocali a livello cross-linguistico.²⁴⁴ Data la ripartizione qui adottata si noterà anche come alcuni studi presenti in Pavlenko (2005) non sono stati inclusi nella presente tabella. Le voci aggiunte e integrate rispetto alla versione di Pavlenko (2005) sono qui state segnalate con un asterisco (*).²⁴⁵

²⁴³ I valori percentuali complessivi sono in questo caso stati calcolati dagli autori solo sulla base delle espressioni emotive, non includendo nel conteggio le produzioni neutre.

²⁴⁴ Ovvero di “studi in cui viene indagata la capacità degli esseri umani a decodificare il parlato emotivo in una o più lingue diverse dalla propria”.

²⁴⁵ L’elenco riportato e integrato potrebbe non essere completo e non si esclude che nella presente indagine possano essere sfuggiti lavori o ricerche che, data la multi- e inter-disciplinarietà dell’argomento trattato, possano trovarsi in riviste di settore che non sono state consultate.

Study	Stimuli	Expression(s)	Languages	Emotions	Results
Kramer (1964)	Content-filtered and unfiltered recordings by 7 male English speakers and 3 Japanese speakers	"There is no other answer. You've asked me that before. My reply has always been the same, and it will always be the same."	<i>Stimulus:</i> English, Japanese <i>Listeners:</i> English	anger, love, contempt, indifference, grief	Same language (unfiltered) 70% Same language (filtered) 61% Foreign language 58%
Beier, Zautra (1972)	unfiltered recordings	(1) "Hello" (2) "Good morning" (3) "How are you?" (4) Random sentence	<i>Stimulus:</i> English <i>Listeners:</i> English, Polish, Japanese	anger, fear, happiness, sadness, indifference, flirtatiousness	same language 60-80% foreign language: Polish 53% Japanese 48%
Solomon, Ali (1975)	unfiltered recordings in English made by one amateur actress	teacher's verbal evaluations in three categories: (1) positive (e.g., "Excellent"); (2) neutral (e.g., "I see"); (3) negative (e.g., "Pretty bad")	<i>Stimulus:</i> English <i>Listeners:</i> Urdu, Hindi, Telugu, Punjabi	positive affect, indifference, negative affect	L1 speakers of English relied more on intonation for affective judgment L2 speakers of English relied more on verbal content
McCluskey, Albas, Niemi, Cuevas, Ferrer (1975)	content-filtered recordings by 3 Canadian and 3 Mexican actresses	two sentences that came first to the speaker's mind to illustrate a particular emotion state	<i>Stimulus:</i> English, Spanish <i>Listeners:</i> English, Spanish	happiness, sadness, love, anger	Mexican boys did better than Canadian ones in identifying emotions both in Spanish and in English
Albas, McCluskey, Albas (1976)	content-filtered recordings by 6 men in each language	two sentences that came first to the speaker's mind to illustrate a particular emotion state	<i>Stimulus:</i> English, Cree <i>Listeners:</i> English, Cree	happiness, sadness, love, anger	each group did better in their own language than in the other language ($p < .01$)
McCluskey, Albas (1981)	content-filtered recordings by 3 women in each language	two sentences that came first to the speaker's mind to illustrate a particular emotion state	<i>Stimulus:</i> English, Spanish <i>Listeners:</i> English, Spanish	happiness, sadness, love, anger	both groups identified emotions better in Spanish; Spanish-speaking participants were better at identification of emotion in both languages

(follows next page)

Study	Stimuli	Expression(s)	Languages	Emotions	Results
Van Bezooijen, Otto, Henan (1983)	unfiltered recordings by 4 males and 4 females	utterance "two months pregnant" in Dutch	<i>Stimulus:</i> Dutch <i>Listeners:</i> Dutch, Japanese, Chinese	anger, shame, sadness, fear, disgust, joy, surprise, interest, contempt, neutral voice	same language 66% foreign language: Chinese speakers 37% Japanese speakers 33%
Magno Caldognetto, Kori (1983)*	unfiltered recordings by 1 Italian professional actor and 1 Japanese non-professional theater actress	three vowels /a/, /e/, /o/	<i>Stimulus:</i> Italian, Japanese <i>Listeners:</i> Italian, Japanese	joy, fear, surprise, sadness, disgust, anger	same language: Italian speakers 87% Japanese speakers 77% foreign language: Italian speakers 53% Japanese speakers 81%
Rintell (1984)	unfiltered recordings	11 conversations where speakers attempted to communicate particular emotions	<i>Stimulus:</i> English <i>Listeners:</i> Arabic, Chinese, Spanish	pleasure, anger, guilt, anxiety, depression, disgust	same language 76% foreign language 33%
Roseberry, McKibbin, Brice (1999)	unfiltered recordings by 2 native speakers of English (male and female) and 2 native speakers of Spanish (male and female)	"Will you bring me the ball?"	<i>Stimulus:</i> English, Spanish <i>Listeners:</i> L1 Spanish, L2 English	happiness, sadness, anger	more correct responses in Spanish than in English ($F(1,8) = 6.96, p < .05$)
Tickle (1999, 2000)*	Unfiltered recordings by 3 native speakers of English and 3 native speakers of Japanese	Non-sense utterances phonotactically possible both in English and Japanese	<i>Stimulus:</i> English, Japanese <i>Listeners:</i> English, Japanese	happy, sad, angry, fearful e calm (as neutral)	same language: English speakers 60% Japanese speakers 36% foreign language: English speakers 35% Japanese speakers 42%
Erickson, Maekawa (2001)	unfiltered recordings	"That's wonderful"	<i>Stimulus:</i> English <i>Listeners:</i> English, Japanese	admiration, suspicion, anger, disappointment	same language 90-100% foreign language 64-92%

(follows next page)

Study	Stimuli	Expression(s)	Languages	Emotions	Results
Scherer, Banse, Wallbott (2001)	recordings by 2 male and 2 female professional German actors	meaningless sentences containing syllables from the six European languages in the study	<i>Stimulus:</i> language-neutral <i>Listeners:</i> German, French, English, Dutch, Italian, Spanish, Indonesian	anger, fear, sadness, joy/happiness, neutral voice (disgust dropped from final analysis due to confusions with anger and sadness)	same language 74% foreign language: Swiss French speakers 69% British & US English speakers 68% Dutch speakers 68% Italian speakers 67% French speakers 66% Spanish speakers 62% Indonesian speakers 52%
Graham, Hamblin, Feldstein (2001)	unfiltered recordings by 2 male and 2 female professional American actors	116 word monologue	<i>Stimulus:</i> English <i>Listeners:</i> English, Japanese, Spanish	anger, fear, joy, sadness, depression, hate, nervousness, neutral voice	same language 59% foreign language: Spanish speakers 42% Japanese speakers 38%
Nakamichi, Jogan, Usami, Erickson (2002)	unfiltered recordings from two versions of one movie	sentences from "Harry Potter and the Sorcerer's Stone"	<i>Stimulus:</i> English, Japanese <i>Listeners:</i> English, Japanese	anger, doubt, surprise, sarcasm, anxiety	same language 91 % foreign language 72 %
Shigeno (1998, 2003)*	vocal and facial expressions in Japanese and American English by 1 Japanese and 1 American actor	five meaningful words or short sentences in both languages	<i>Stimulus:</i> Japanese, American <i>Listeners:</i> Japanese, Nord American	happiness, surprise, anger, disgust, fear, and sadness	percent of correct responses of the six emotions ranged from 64% to 86% for the auditory-alone condition: Japanese subjects recognized the vocal expression of Japanese better than that of English, while the Americans recognized English vocal expression better than that of Japanese.
Dromey, Silveira, Sandor (2005)*	unfiltered recordings by 1 English Mother Tongue (EMT) professional actor	A list of 234 single words (from 1 to 3 syllables)	<i>Stimulus:</i> english <i>Listeners:</i> monoglot English Mother Tongue, poliglott English Mother Tongue, Other Mother Tongue	anger, neutral	Higher recognition rates for poliglott EMT listeners, followed by Other Mother Tongue and monoglot EMT listeners

(follows next page)

Study	Stimuli	Expression(s)	Languages	Emotions	Results
Sawamura <i>et al.</i> (2007)*	unfiltered recordings from an emotional speech database produced by 15 Japanese children	The word “Pikachu” without linguistic information from the famous animation of “Pocket Monster”	<i>Stimulus:</i> language-neutral <i>Listeners:</i> Japanese, American, Chinese	anger, joy, sad and surprise	mean accuracy rate ca 60%
Shochi, Aubergé, Rilliard (2007)*	unfiltered recordings produced with 12 attitudes by 1 Japanese subject	7 sentences (1, 2, 5 or 8 moras) with either a single word or a verb-object structure as the syntactic structure	<i>Stimulus:</i> Japanese <i>Listeners:</i> Japanese, French, American English	doubt-incredulity (DO), evidence (EV), exclamation of surprise (SU), authority (AU), irritation (IR), arrogance (AR), sincerity-politeness (SIN), admiration (AD), kyoshuku (KYO), simple-politeness (PO), declaration (DC) and interrogation (IN)	according to a chi-square test, the distribution of attitudes is above chance for main attitudes with higher recognition rates for Japanese listeners and some differences in western listeners
Pell, Monetta, Paulmann, Kotz (2009)*	“pseudo-utterances” produced by native speakers of four different languages: (Argentine) Spanish, (Canadian) English, (Standard high) German, and (Jordanian/Syrian) Arabic.	pseudo-utterances retaining natural phonological and morphosyntactic properties of the target language with meaningful content words (e.g., nouns, verbs) replaced with plausible pseudo-words (e.g., for English: The fector egzullin the boshent)	<i>Stimulus:</i> Spanish, English, German, Arabic. <i>Listeners:</i> monolingual, Spanish-speaking adults	anger, disgust, fear, sadness, joy, pleasant surprise and neutral	same language 64% foreign language: Arabic 59% English 58% German 56%

Tabella 14: Studi di tipo cross-linguistico volti al *decoding* delle emozioni trasmesse attraverso il canale vocale (integrazione ed adattamento da Pavlenko, 2005, pp. 58-61, table 3.2.)

Il dato più significativo che emerge dagli studi sin qui riportati, e riassunti brevemente in Tabella 14, è sicuramente attribuibile al fatto che gli esseri umani sono in grado di decodificare correttamente produzioni vocali emotive con percentuali che vanno ben oltre la pura casualità, anche nel caso in cui si affronti la tematica del parlato emotivo in lingue diverse dalla propria, come quella presente in contesti cross-linguistici e cross-culturali.

Dagli studi appena menzionati emerge, inoltre, come l'aspetto del *decoding* cross-linguistico e cross-culturale sia stato investigato in varie forme: ad esempio prendendo in esame contesti culturali nettamente contrapposti e spesso agli antipodi (come americani vs. giapponesi o ancora cinesi); utilizzando stimoli vocali composti da frasi *meaningless*, semplici parole (sia in versione originale che in versione filtrata e/o modificata), stimoli che in ciascun esperimento non sono stati mai raccolti in più di tre lingue. Pochissimi tra questi, infine, si sono occupati esclusivamente di lingue europee e ancora meno sono gli studi su stimoli in lingua italiana o su ascoltatori italiani.

Si tratta, tutto sommato, di semplici considerazioni che sono però funzionali al proseguimento della presente esposizione. Tali considerazioni costituiscono, inoltre, una base di partenza per quella che è la ricerca qui intrapresa e in base alle quali molte delle scelte metodologiche a cui si accennerà più avanti risulteranno motivate oltre che determinate.

7.3 *Encoding delle emozioni a livello cross-linguistico*

Di numero decisamente inferiore rispetto agli studi relativi al *decoding* cross-linguistico delle emozioni vocali sono quelli rivolti all'*encoding*.

In Piot (1999) viene presentato un esperimento sull'uso di elementi prosodici che persone con diverso background culturale utilizzano per esprimere emozioni o per conferire specifiche attitudini emotive al parlato. L'esperimento coinvolge 4 lingue (anglo-americano, coreano, finlandese, vietnamita) con un unico corpus contenente due frasi²⁴⁶ registrato da quattro soggetti (due di sesso maschile e due di sesso femminile) di ciascuna lingua con tredici diverse attitudini emotive.²⁴⁷ In base alle analisi acustiche condotte per ciascuna delle tredici attitudini emotive nelle quattro lingue, e fermo restando le differenti caratteristiche prosodiche delle lingue esaminate, secondo l'autore, che evidenzia la necessità di analisi più approfondite, emergerebbero similitudini significative nell'espressione di emozioni ed attitudini.

²⁴⁶ Una frase contenente un dislocamento del soggetto e una frase composta da una singola parola, ovvero un nome ("Natasha").

²⁴⁷ Per i dettagli sulla creazione e la composizione del corpus si rimanda a Piot (1999), pag. 369.

A completamento di una serie di ricerche precedenti²⁴⁸, Kori, Magno Caldognetto (2003) hanno condotto uno studio contrastivo tra parlato emotivo italiano e parlato emotivo giapponese al fine di verificare:

- se i tratti fonetici macroprosodici attraverso cui le emozioni vengono veicolate costituiscano un sistema di opposizioni tra *patterns* discreti (specifici di ciascuna lingua e cultura) o se, al contrario, tali tratti siano simili nelle due lingue esaminate;
- come nelle realizzazioni emotive vengano mantenute o modificate le caratteristiche acustiche che determinano l'accento lessicale nelle due lingue in esame (accento dinamico, di intensità, per l'italiano, musicale per il giapponese).

Per rispondere ai quesiti che i due autori si sono posti, sono state utilizzate due frasi composte da un sintagma nominale e uno verbale (SN + SV): per l'italiano /la'mam:a'mandʒa/ [la'mam:a'mandʒa] ("la mamma mangia"); per il giapponese /mamaga najanderu/ [mamaŋa najanderu] ("la mamma sta pensando"). Le registrazioni, effettuate in cabina silente, sono state prodotte, sulla base di appositi scenari, rispettivamente da un attore italiano e da uno *speaker* di una rete nazionale giapponese in sei stati emotivi primari e con intonazione neutra (intonazione modale dichiarativa). Ciascuna produzione è stata ripetuta per almeno cinque volte. Le produzioni sono successivamente state sottoposte a test di adeguatezza con un gruppo di sei giudici esperti che ne ha valutato l'adeguatezza su una scala a 5 punti: sono state quindi ritenute e analizzate acusticamente le tre produzioni più adeguate. Per l'analisi acustica sono state analizzate le durate dei singoli sintagmi, i valori di frequenza fondamentale²⁴⁹, spettrogrammi a banda larga e spettri medi delle singole vocali. Le analisi macroprosodiche condotte hanno rivelato da un lato una serie di parallelismi, dall'altro una serie di differenze tra i dati relativi all'italiano e quelli relativi al giapponese. In merito alla durata il *range* di variazione nel giapponese è risultato assai ridotto rispetto all'italiano in cui sono stati riscontrati valori assoluti nettamente superiori e molto diversi tra loro: nel giapponese non è stato, quindi, rilevato un ricorso a variazioni di durata per differenziare tra loro le emozioni. Per le variazioni di intensità è stato, invece, rilevato un ricorso sistematico e parallelo a questo indice acustico da parte di entrambi i parlatori con i valori più elevati per

²⁴⁸ Si veda quanto riportato nel paragrafo dedicato al *decoding* cross-linguistico, nello specifico Magno Caldognetto, Kori (1983).

²⁴⁹ In questo caso tutte le analisi relative alla F_0 sono state effettuate sulla base dei rispettivi valori in semitoni calcolati secondo la formula $12 \cdot \log(F_0 \text{ in Hz}) / \log(50)$.

collera e i più bassi per tristezza e disgusto. Nel caso di F_0 media è stata rilevata una buona coerenza in entrambe le lingue: valori alti per collera, gioia e sorpresa e valori più bassi registrati per tristezza. Nonostante ciò, però, gli autori sottolineano, in questo caso, una serie di eccezioni: nel caso del parlante italiano l'emozione di paura è stata realizzata con i valori di F_0 più alti rispetto a tutte le altre emozioni, mentre molto vicini a quelli di tristezza sono stati riscontrati quelli per il disgusto. Al contrario, nel giapponese queste due emozioni sono state realizzate con valori di F_0 media collocati nella zona centrale.²⁵⁰ Per l'intensità e l' F_0 è stata, inoltre, rilevata una correlazione sulla base delle previsioni delle caratteristiche acustiche che veicolano le emozioni elaborate in base alle dimensioni psicologiche che specificano le emozioni primarie (attivazione, valutazione e controllo): è stata verificata l'opposizione tra emozioni ad alta attivazione (collera, gioia) ed emozioni a bassa attivazione (tristezza).

Sempre voci asiatiche sono le voci confrontate con voci italiane in Anolli *et al.* (2008a, 2008b). In questo studio sono state acquisite produzioni emotive prodotte da 29 giovani adulti cinesi e 19 giovani adulti (ciascuno nella propria lingua) ricavate dalla lettura di testi emotigeni relativi a 8 emozioni (gioia, tristezza, collera, paura, disprezzo, orgoglio, colpa), più una modalità neutra, all'interno dei quali era stata inserita una frase standard.²⁵¹ Anche in questo caso le analisi acustiche condotte sulle produzioni acquisite sono quelle relative al tempo, alla frequenza fondamentale e all'intensità.²⁵² Dai risultati delle analisi acustiche e statistiche condotte è emerso che l'espressione delle emozioni vocali nella lingua italiana e cinese non sembra essere identica, con una sostanziale differenza soprattutto rispetto alle variazioni delle espressioni emotive in base alla produzione neutra, maggiore nella cultura italiana, molto minore nella cultura cinese. È stata anche rilevata una minore distinzione nei soggetti cinesi tra emozioni piacevoli ed emozioni spiacevoli, distinzione che nei soggetti italiani risulta molto più marcata, confermando così la presenza di modelli culturali

²⁵⁰ La spiegazione fornita dagli autori viene in questo caso ricondotta alla possibile presenza di specifiche norme di esibizione culturale anche a causa dei diversi contenuti semantici e relative etichette lessicali utilizzate per l'elicitazione delle produzioni emotive ai parlanti.

²⁵¹ La frase era in italiano "non è possibile, non ora" mentre in cinese (mandarino) "bu keneng, zhe bu keneng" (letteralmente "non è possibile, questo non è possibile"). Sui testi emotigeni si è già ampiamente dibattuto nei precedenti paragrafi.

²⁵² Non è stata, tuttavia, effettuata alcuna validazione del materiale sonoro acquisito, probabilmente per il fatto che era stata operata una validazione dei testi emotigeni da parte di un gruppo di 30 giudici che hanno giudicato il riconoscimento univoco dell'emozione, la validità interna e di costruito delle storie ecc..

diversificati nelle due lingue.²⁵³ Alla luce di quanto detto, i risultati contribuirebbero secondo gli autori,

*“a comprendere come i modelli culturali influenzino le modalità vocali standard nell’espressione delle emozioni. I modelli espressivi prevalenti tendono a sintonizzarsi con la prospettiva cultura specifica (cinese o italiana) e le configurazioni vocali caratteristiche delle emozioni possono essere considerate come un prodotto dei modelli culturali focali”.*²⁵⁴

Sulla base dei risultati è emerso come, nello specifico, diverse emozioni possono essere trasmesse in modo discreto attraverso la produzione vocale in entrambe le culture, consentendo anche l’identificazione di modelli culturali distintivi di indici vocali nell’espressione delle emozioni. Gli autori, ribadendo lo spirito esplorativo dello studio, riconoscono anche una serie di limiti della ricerca da loro condotta. Tra questi vengono citati:

- possibili artefatti legati alla differenza di struttura linguistica delle due lingue studiate;
- restrizione del confronto a sole due lingue, di cui una di cultura orientale ed una di cultura occidentale;
- utilizzo di studenti universitari in qualità di attori-lettori “ingenui” per l’espressione delle emozioni.

Al di là delle limitazioni, seppur presenti, gli autori fanno rilevare anche una serie di contributi alla comprensione della produzione interculturale delle emozioni. Lo studio condotto e i risultati raccolti sembrano, infatti:

- essere coerenti con il modello di “controllo di valutazione dello stimolo” (*stimulus evaluation check – SEC*) proposto da Scherer (1984, 1986, 2003; Banse, Scherer, 1996)²⁵⁵;
- supportare l’ipotesi evuzionista della comunicazione delle emozioni che vedrebbe il sistema vocale come il canale non verbale dotato di maggiore continuità filogenetica.²⁵⁶

²⁵³ Questa sostanziale differenza viene ricondotta dagli autori al fatto che i cinesi enfatizzano molto di più l’armonia sociale che li limita dall’occupare troppo spazio nelle relazioni sociali con gli altri individui. Si parla, infatti, di società collettivistica per i cinesi e di società individualista nelle culture occidentali. Per maggiori dettagli si rimanda ad Anolli *et al.* (2008), pp. 34-35.

²⁵⁴ Cfr. Anolli *et al.* (2008a), pag. 36.

²⁵⁵ Stando alle predizioni formulate da Scherer a proposito degli indici acustici associati alle diverse emozioni, in cui le emozioni coinvolgono un sistema sequenziale di valutazioni cognitive (cfr. quando detto più sopra nel § 4.2), il prodotto di ogni controllo di valutazione dello stimolo (*SEC*) produce un particolare effetto sul sistema nervoso e, di conseguenza, influenza la muscolatura responsabile della produzione vocale.

²⁵⁶ Cfr. Anolli *et al.* (2008a), pag. 33.

Un singolare esperimento condotto da Braun, Oba (2007), ha invece affrontato l'aspetto interculturale delle emozioni base (*anger, joy, fear, sadness*, come anche di produzioni verbali neutre) utilizzando campioni di parlato doppiato prendendo in esame alcune scene di una popolare serie TV americana (*Ally McBeal*) e le rispettive parti doppiate in lingua tedesca e giapponese. Dalle scene tratte sono state analizzate le produzioni di due protagonisti della serie TV (uno maschile e uno femminile) con riferimento a processi di *encoding* dei vari aspetti di *speaking tempo*. Sebbene la discussione dei risultati da parte di Braun, Oba resta di enorme difficoltà per l'assenza, come loro stessi affermano, di studi simili, i risultati, che si concentrano maggiormente su questioni legate ai vari aspetti di *speaking tempo*, provano la presenza di maggiori differenze culturali e differenze di sesso nel processo di *encoding*. Differenze tra i soggetti femminili risultano essere maggiori tra lingue linguisticamente e culturalmente meno vicine, mentre differenze tra i soggetti maschili non sono così evidenti. Sulla base dei loro risultati non viene, invece, confermata la conclusione a cui sono giunti Pettorino, Vitagliano (2003). Secondo questi ultimi gli adattamenti temporali nel doppiaggio vengono primariamente effettuati attraverso l'adattamento della lunghezza delle pause con un allungamento delle stesse nella versione doppiata, se il testo dell'originale risultava più lungo e viceversa, ovvero un aggiustamento (riduzione) delle pause piuttosto che un intervento sullo *speech rate*. Analogamente, alcuni dei risultati frutto di precedenti ricerche come anche alcune delle predizioni formulate da Banse, Scherer (1996) non vengono completamente confermati: ad esempio per la paura, che evidenzia, nel caso in oggetto, un incremento dell'*articulation rate* solo in due delle lingue esaminate rimanendo del tutto inalterato nel tedesco.

7.4 *Encoding/decoding delle emozioni a livello cross-linguistico*

Se, come visto sino ad ora, il numero degli studi può essere considerato di tipo decrescente passando da studi volti al *decoding* a studi volti all'*encoding*, di numero ancora più esiguo sono quegli studi che hanno affrontato il duplice aspetto dell'*encoding/decoding* delle emozioni espresse a livello orale.

Chung (1999, 2000) si è, ad esempio, interrogata su come l'*arousal* emotivo modifichi la produzione vocale nel parlato e come questa venga percepita da soggetti ascoltatori in base a fattori linguistici, acustici e culturali diversi con possibili applicazioni nei campi dello *speech synthesis* e dello *speech recognition*. Il materiale relativo al parlato emotivo, in questo caso parlato spontaneo e naturale, è stato ricavato da registrazioni di interviste televisive di parlanti femminili di lingua coreana e americana. All'interno delle interviste utilizzate sono

state identificate, per ciascuna parlante, distinte transizioni emotive: una parte iniziale, in cui l'intervistata è gioiosa e sorridente e in cui parla di momenti felici trascorsi con i propri familiari (considerata per questo come gioia); una parte centrale dell'intervista in cui la donna parla tranquillamente senza particolari espressioni emotive (considerato come momento rappresentante la sua voce neutra) e, infine, una parte terminale in cui l'intervistata inizia a parlare di problemi e questioni conflittuali con la propria famiglia divenendo triste e finendo per singhiozzare dal pianto (considerata come tristezza). Sono state quindi condotte tre diverse tipologie di analisi: analisi acustiche, esperimenti percettivi ed analisi condotte per mezzo di procedure di sintesi vocale. Per quel che concerne le analisi acustiche condotte sul materiale in coreano è emerso che l'*arousal* emotivo (nei confronti di gioia o di tristezza) produce un incremento dei valori massimi di F_0 e del suo *range*: nella gioia è stato riscontrato un innalzamento della F_0 media, mentre nella tristezza un decremento nella F_0 media e nei valori inferiori di F_0 . Per quel che riguarda, invece, *jitter* e *shimmer*, i valori sembrano subire un innalzamento sotto l'effetto dell'*arousal* emotivo, mentre lo *speaking rate* sembra aumentare con la gioia e diminuire con la tristezza (anche se questi ultimi dati non fanno rilevare una significatività statistica). Il test percettivo condotto con soggetti coreani, francesi e americani conferma la presenza nella comunicazione orale delle emozioni di aspetti sia universali che cultura-specifici: i soggetti dei tre gruppi di ascoltatori hanno, infatti, identificato le emozioni di gioia e di tristezza espresse da un soggetto coreano con un grado di accuratezza superiore al caso. Anche in questa occasione, come già in molti altri studi qui citati, è emersa una maggiore precisione e abilità nei riconoscimenti corretti da parte di soggetti (coreani) aventi la stessa lingua degli stimoli presentati (coreano), rivelando ancora una volta la presenza di un *in-group advantage*. I risultati conseguiti nelle due fasi appena riportate sono poi stati verificati attraverso due ulteriori esperimenti percettivi con ascoltatori coreani e francesi ai quali sono stati presentati stimoli in cui, attraverso procedimenti di sintesi vocale, erano stati modificati i parametri relativi ai contorni di F_0 e di durata: in questo caso i risultati hanno fatto evidenziare come proprio la parte finale della frase sia maggiormente efficace rispetto alla parte iniziale e centrale quando si tratti di comunicare le emozioni.

Affrontando la questione da un punto di vista più "europeista", Abelin. Allwood (2000, 2002) si sono, invece, chiesti se:

1. vi fosse stabilità nel modo in cui interpretiamo diverse emozioni e attitudini a partire da *patterns* prosodici;

2. se questa interpretazione sia dipendente dal background linguistico e culturale dell'ascoltatore;
3. se vi sia qualche relazione ricorrente tra proprietà acustiche e semantiche dello stimolo.

Per rispondere ai quesiti appena riportati è stata registrata la produzione di una frase²⁵⁷ con diversi contenuti emotivi da parte di un soggetto svedese. Il materiale è stato successivamente sottoposto all'ascolto di soggetti con diversa L1 (svedese, inglese, finlandese e spagnolo), con il compito di valutare il contenuto emotivo delle espressioni scrivendo su un foglio di carta lo stato emotivo che ad esse attribuivano.²⁵⁸ Tutte le risposte fornite sono state raccolte e organizzate in campi semantici anche se, in alcuni casi, gli autori hanno avuto difficoltà nella classificazione, difficoltà dovute o alla fantasia degli ascoltatori, o alla mancanza di corrispondenze cross-linguistiche univoche delle etichette, sollevando, qui, una questione di cruciale importanza nelle ricerche sulle emozioni a livello cross linguistico:

"It is obvious that emotions are expressed (also non verbally) in different ways in different cultures, but are also certain emotions experienced in certain cultures but not in others?"²⁵⁹

I risultati hanno rivelato che alcune emozioni vengono interpretate in accordo con le emozioni intese più di altre (*anger, fear, sadness e surprise*), mentre altre in misura inferiore. Inoltre le emozioni vengono interpretate con diversi gradi di successo in base alla L1 degli ascoltatori. Gli ascoltatori nativi hanno naturalmente i risultati migliori. Riguardo, invece, alla relazione tra proprietà acustiche e semantiche dello stimolo di cui al punto tre, sono state rilevate, a livello acustico, similarità tra alcune emozioni (come ad esempio *fear e dominance*, che si rassomigliano molto l'una all'altra per durate brevi ed forte intensità).²⁶⁰

Breitenstein, Van Lancker, Daum (2001) si sono interrogati sull'utilizzo degli indici acustici nella percezione di emozioni vocali da parte di soggetti ascoltatori. Nello specifico gli autori hanno presentato due set di stimoli vocali emotivi, prodotti in lingua tedesca da un'attrice, a due gruppi di ascoltatori composti da 35 tedeschi e 30 americani. I due set di stimoli vocali erano composti da: frasi in lingua tedesca pronunciate con cinque diverse emozioni vocali (*happy, sad, angry, frightened e neutral*); una versione degli stimoli emotivi

²⁵⁷ È stata scelta una frase (*carrier phrase*) con termini relativi al cibo, in quanto, secondo gli autori, si possono esprimere emozioni attraverso termini legati al cibo mantenendone il contenuto neutro rispetto alle emozioni.

²⁵⁸ Dove il termine dello stato emotivo attribuito veniva riportato nella lingua madre si è successivamente provveduto alla sua traduzioni in svedese.

²⁵⁹ Cfr. Abelin, Allwood (2000), pag. 111.

²⁶⁰ Riguardo quest'ultimo aspetto anche in questo caso i dati sono, per le motivazioni a cui più sopra si accennava, alquanto carenti.

originali con alterazione o manipolazione sistematica dei parametri relativi a F_0 (*pitch*) e durata (*rate*). L'istruzione data agli ascoltatori prevedeva una duplice valutazione degli stimoli presentati in base alle cinque categorie emotive proposte e in base alla percezione del grado di attivazione dello stimolo su una scala a cinque punti riportante alle due estremità i valori *very active* e *very passive*.²⁶¹ Riguardo gli stimoli vocali non-manipolati sono stati riscontrati valori di riconoscimento in linea con quelli di precedenti studi condotti da Banse, Scherer (1996) e Scherer *et al.* (1991) con percentuali di riconoscimento corretto maggiori per le categorie emotive di *anger* e *sadness* e minori per *happiness*. Per quel che concerne gli stimoli manipolati, invece, uno *slow rate* negli stimoli presentati è stato associato all'etichetta *sad*, mentre un *fast rate* è stato classificato come *angry*, *frightened* o *neutral*. Rispetto alla manipolazione degli indici di durata degli stimoli, l'alterazione del *pitch* ha influenzato in misura minore la scelta delle categorie emotive vocali: un riduzione della variazione del *pitch* è stata associata alla percezione di *sad* o *neutral*; una maggiore variazione del *pitch* ha prodotto, invece, maggiori risposte a favore delle emozioni di *frightened*, *angry* e *happy*. Nel complesso la performance dei due gruppi di ascoltatori si è rivelata assai simile, sebbene, come fanno notare gli autori, gli ascoltatori tedeschi hanno percepito maggiore variabilità nel grado di attivazione degli stimoli emotivi rispetto agli ascoltatori americani.²⁶² L'esperimento condotto ha in buona sostanza evidenziato come soggetti ascoltatori siano in grado di percepire le sottostanti emozioni in stimoli vocali anche nel caso di alterazioni acustiche volontarie²⁶³ e come indici vocali aggiuntivi siano effettivamente importanti per l'ascoltatore.

Burkhardt *et al.* (2006) hanno, invece, condotto un esperimento comparativo su più lingue (francese, tedesco, greco e turco) allo scopo di studiare gli effetti percettivi di modificazioni prosodiche in stimoli di parlato emotivo sintetizzati. Sette frasi semanticamente simili ed emotivamente caratterizzate (*neutral*, *joyful*, *friendly*, *threatening*, *bored*, *frightened*, *sad*) sono state tradotte nelle lingue in esame per poi essere sintetizzate e sistematicamente manipolate con riferimento a *pitch range*, durata e *jitter* con l'ausilio di un apposito *software*.

²⁶¹ La scelta di inserire una scala di attivazione viene motivata dagli autori dal fatto che: a) differenze nel *pitch range* hanno effetti continui (piuttosto che categorici) sulle valutazioni degli stimoli vocali emotivi da parte dell'ascoltatore (vedi Ladd *et al.*, 1985), particolarmente nelle attribuzioni di *arousal*; b) la dimensione di attività è correlata con più ampie variabilità del *pitch* e *speech rate* più elevato (vedi Frick, 1985).

²⁶² Rispetto a precedenti studi (ad es. Albas *et al.*, 1976; Van Bezooijen *et al.*, 1983) le risposte non sono risultate chiaramente correlate alla dimensione di attivazione. In aggiunta a ciò gli autori fanno anche rilevare come da parte degli ascoltatori è stata spesso scelta la categoria neutra quando non corretta. Poiché in precedenti studi condotti in ambito cross-culturale (gli autori citano in questo caso Kramer, 1964 e Beier, Zautra, 1972, ma vale ovviamente per la maggior parte degli studi che sino a qui sono stati riportati) si rileva la totale assenza di una categoria neutra, o comunque non emotiva, la possibilità di valutare e confrontare i risultati emersi in questo e in altri studi risulta, fanno notare Breitenstein *et al.* (2001), assai limitata.

²⁶³ Gli autori fanno tuttavia notare come si tratti, in questo caso, di stimoli vocali sintetizzati di ottima qualità.

Gli esperimenti percettivi somministrati a gruppi di soggetti ascoltatori dei paesi esaminati hanno rivelato effetti significativi per le variabili manipolate indipendentemente dalla lingua presentata, dal sesso e dall'età dei soggetti.

Thompson, Balkwill (2006) si sono recentemente occupati del riconoscimento di emozioni vocali (*joy, anger, sadness e fear*)²⁶⁴ prodotte in lingua inglese, tedesca, cinese, giapponese e tagalog, da parte di venti soggetti ascoltatori di madrelingua inglese ricercando nelle proprietà acustiche degli stimoli presentati²⁶⁵ elementi di correlazione con le emozioni codificate. L'accuratezza del corretto riconoscimento delle produzioni emotive è stata al di sopra del caso per tutte le lingue presentate agli ascoltatori inglesi: in tutte le lingue le emozioni maggiormente identificate sono state *sadness e anger* contro *joy e fear* per le quali sono stati registrati valori di corretta identificazione più bassi. Anche in questo caso è stato registrato un *in-group advantage*, in cui sono state riconosciute in misura maggiore le produzioni nella lingua madre degli ascoltatori (94%) rispetto a quelle relative al giapponese e al cinese che sono state identificate in modo corretto in misura inferiore (rispettivamente del 54% e 59%). Per quanto concerne, invece, le proprietà acustiche degli stimoli, confermandone la variazione in funzione delle emozioni espresse, è risultata significativa la variazione della media di F_0 e del *range* in funzione dell'emozione intesa: rispettivamente medie più alte associate alle emozioni di *joy e anger*, piuttosto che con *fear e sadness* e valori relativi al *range* di F_0 più alti in *joy* e più bassi in *sadness*. Analoghe considerazioni fatte per F_0 media valgono per intensità media. Unica variazione significativa tra le lingue è stata rilevata dall'associazione tra *range* dell'intensità ed emozione. Thompson, Balkwill (2006) concludono supportando l'idea che la prosodia emotiva viene decodificata sulla base tanto di indici universali quanto di indici cultura specifici.

7.5 Conclusioni

In conclusione si può sicuramente affermare come, dalla panoramica di studi riportati e dai risultati a cui i vari autori sono giunti, le emozioni vocali siano, alla pari di quelle facciali, riconosciute cross-linguisticamente e cross-culturalmente, con risultati nettamente al di sopra della semplice casualità.

²⁶⁴ Le frasi registrate del tipo "*the bottle is on the table, the leaves are changing color*" sono state pronunciate nelle rispettive lingue con le quattro emozioni di *joy, anger, sadness e fear*. Manca in questo caso una produzione neutra.

²⁶⁵ Stimoli che erano stati preventivamente validati in ciascuna lingua da due soggetti ascoltatori allo scopo di selezionare solo quelle produzioni che fossero "[...] *exemplary of how prosody is typically used in each language*", escludendo in questo modo "[...] *exaggerated or dramatic uses of prosody*" (cfr. Thompson, Balkwill, 2006: 411).

Inoltre, sebbene si sia tentato in queste pagine di fornire un quadro ampio e il più possibile dettagliato sullo stato dell'arte dell'argomento trattato, un'analisi comparativa dei risultati resta sempre assai complessa e delicata. Tuttavia, qualora si decidesse di effettuare un simile resoconto, come per altro molti hanno già fatto (vedi ad es. Elfenbein, Ambady, 2002; Juslin, Laukka, 2003), c'è indubbiamente da tener conto di una serie di limitazioni e problematiche connesse, sulle quali già Magno Caldognetto, Kori (1983: 360)²⁶⁶ avevano posto l'accento:

“The scores of correct identification as well as the patterns of confusion, depend on the quality of the test stimuli and on the number of categories of emotion, it would be unsafe to compare the results of different studies or to generalize particular observations based on the identification test without any assessment of the quality of the stimuli.”

Terminata questa ampia rassegna di studi si passerà ora alla definizione di quelli che sono gli obiettivi a cui si tende con la ricerca intrapresa.

²⁶⁶ Ma anche Kramer (1964) che, ad esempio, evidenziava la stretta relazione dei risultati del suo test di identificazione con la rappresentatività degli stimoli utilizzati nello stesso test, un'osservazione sicuramente estendibile a tutti gli studi sino ad ora condotti.

8 Ipotesi di lavoro e obiettivi della ricerca

La capacità dell'uomo a riconoscere emozioni espresse attraverso la voce è stata ampiamente documentata in letteratura. Altrettanto condiviso è il fatto che, come esseri umani, siamo in grado di riconoscere le emozioni ad un livello più alto, quale quello che può essere definito cross-linguistico e cross-culturale.

Come già ampiamente emerso dagli studi passati in rassegna nel capitolo 7, se da un lato si rilevano innumerevoli studi sul parlato emotivo, dall'altro gli studi di tipo cross-linguistico e cross-culturale risultano essere assai sparuti sia da un punto di vista di *decoding*, come anche di *encoding* o di entrambi. I motivi di questa sporadicità non sono affatto legati, come già detto, all'assenza di interesse per l'argomento, ma sono prevalentemente imputabili alla difficoltà che lo studio delle emozioni vocali impone, difficoltà ulteriormente esacerbate nello studio di tipo cross-linguistico-culturale.

Dai risultati degli studi sino ad oggi condotti emerge, infatti, come le emozioni vocali, alla pari di quelle facciali, siano riconosciute cross-linguisticamente e cross-culturalmente con risultati nettamente al di sopra della semplice casualità. Da una meta-analisi condotta da Laukka (2004) sugli studi presenti in letteratura riguardo alle emozioni vocali, allo scopo di verificare se le emozioni siano riconosciute cross-culturalmente e se vi siano specifici *patterns* acustici della voce per categorie discrete di emozioni, emerge come: a) l'accuratezza di riconoscimento è superiore a quella data dalla casualità per categorie di emozioni più ampie; b) il *decoding* cross-culturale è mediamente inferiore al *decoding* intra-culturale del 7%; c) esistono specifici *patterns* acustici nella voce delle emozioni che vengono utilizzati per comunicare emozioni discrete.

Ciononostante, se da una parte Scherer, Banse, Wallbott (2001: 78) affermano che:

“it seems reasonable to assume that the recognition of vocal emotion expressions might work across language and culture boundaries”,

più avanti gli stessi autori, sottolineano come

“[...] encoders and decoders from several different countries would need to be studied, allowing the construction of an encoder-decoder-emotion matrix and to test whether decoders

*from the countries involved would recognize emotion portrayals by encoders from their own countries most accurately”.*²⁶⁷

È proprio nel quadro qui descritto da Scherer, Banse, Wallbott (2001: 88) e Scherer (2000: 382) che si inserisce il presente lavoro. Partendo, pertanto, da tali presupposti, l'intera ricerca è stata sviluppata in tre distinti *step*, ciascuno dei quali con obiettivi specifici e con funzioni propedeutiche tra il primo e i successivi *step*. In Figura 35 viene fornita una rappresentazione schematica dei tre *step*:

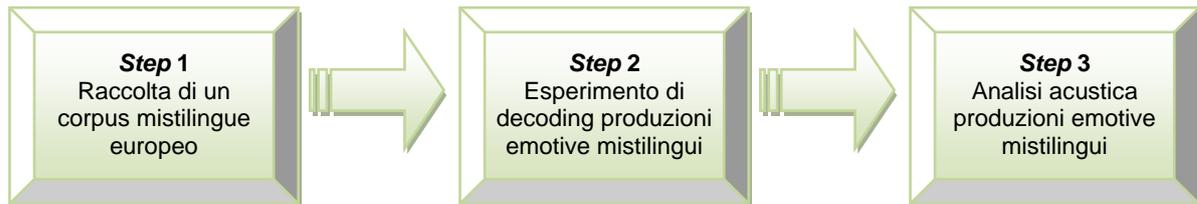


Figura 35: Rappresentazione schematica dei tre *step* della ricerca proposta

Di seguito gli le ipotesi e gli obiettivi alla base di ciascuno degli *step* della ricerca condotta.

8.1 Step 1: raccolta di un corpus mistilingue europeo

Una ricognizione su ben 64 database di parlato emotivo (*cfr.* Ververidis, Kotropoulos, 2006) ha rilevato l'assenza di *corpora* di parlato emotivo mistilingue utili per uno studio cross-linguistico-culturale delle emozioni. Ciò ha reso necessario la raccolta di produzioni verbali emotive nelle quattro lingue (italiano, francese, inglese e tedesco) con riferimento alle emozioni definite da Ekman (1992) come *basic* (*happiness, anger, fear, sadness, disgust, surprise*). Per l'attuazione delle successive fasi della ricerca, in questa fase si è inoltre resa assolutamente necessaria una fase di validazione percettiva del materiale raccolto per la costituzione del corpus di parlato emotivo mistilingue.

Gli obiettivi del primo *step* della ricerca saranno essenzialmente quelli di:

1. motivare e illustrare le caratteristiche del *corpus* di parlato emotivo mistilingue “europeo” raccolto in via sperimentale²⁶⁸ per le lingue italiana, francese, inglese e tedesca, con particolare riferimento al protocollo di elicitazione adottato, al numero e alla tipologia di soggetti registrati;
2. fornire i risultati della procedura di validazione percettiva del corpus raccolto.

²⁶⁷ *Cfr.* Scherer, Banse, Wallbott (2001), pag. 88.

²⁶⁸ Per questo aspetto non è stato possibile rispondere appieno ai *desiderata* di Douglas-Cowie *et al.* (2003).

Con riferimento al primo punto, si illustrerà nei dettagli il protocollo di elicitazione costituito da tre fasi distinte e tra loro conseguenti ispirate al “paradigma degli scenari” e al “contenuto standard” di Anolli *et al.* (2008a, 2008b), Anolli, Ciceri (1992) e Scherer *et al.* (1991) con la raccolta in quattro lingue di produzioni da parte sia di soggetti *naif* che di *attori*.

Con riferimento al secondo punto, invece, verranno presentate le procedure e soprattutto i risultati della fase di validazione del corpus effettuata con la somministrazione a gruppi di ascoltatori di due test di validazione, rispettivamente di:

- a. identificazione (T1), attraverso la verifica di eventuali differenze di riconoscimento da parte dei soggetti ascoltatori coinvolti nelle registrazioni, con riferimento a ciascuna delle tre fasi del protocollo di elicitazione;
- b. rappresentatività delle produzioni emotive (T2), per quantificare e valutare il contributo, in termini di materiale utile e di riconoscibilità delle produzioni, da parte di soggetti *naif* da un lato e di *attori* dall'altro.

L'attuazione di questo *step*, oltre a fornire il materiale utile all'espletamento delle successive fasi della ricerca, avrà come obiettivo anche la produzione di un corpus di parlato emotivo mistilingue europeo raccolto a livello sperimentale che contribuirà a colmare una lacuna caratterizzata dall'assenza di *corpora* mistilingui (che contengano lingue europee, italiano compreso) all'interno del panorama delle risorse di parlato emotivo (*cf.* § 6.6).

8.2 Step 2: decoding di produzioni emotive mistilingui

Sebbene la comunicazione sia da intendere come una continua integrazione di diversi canali di espressione (come ad es. quello del volto, dei gesti e non ultima della voce) siamo anche in grado di utilizzarli singolarmente, come avviene, ad esempio, nel caso estremo di soggetti sordomuti che, non potendo fare ricorso al canale vocale e uditivo, fanno ricorso a quello gestuale e visivo; o ancora a quello di soggetti privi della vista che fanno esclusivo riferimento al canale vocale e uditivo, oltre che a quello tattile. Si tratta naturalmente di casi particolari che ben documentano le abilità a cui in questa sede si fa riferimento, dove determinati canali di comunicazione vengono sviluppati più di altri per sopperire alle carenze o mancanze a carico di determinati organi sensoriali. Sulla scia degli esempi appena presentati, non è necessario andare molto lontano per rendersi conto di quanto accade per la comunicazione verbale in situazioni normali, alla quale ci si rivolgerà in questa sede.

Si è più sopra visto (vedi § 4.1) come, attraverso la comunicazione verbale, trasmettiamo ai nostri interlocutori molteplici informazioni che Magno Caldognetto *et al.* (2008) hanno definito come informazioni di tipo linguistico, para-linguistico ed extra-

linguistico. Ad esempio, nella realtà quotidiana e nel caso di soggetti normodotati, in assenza di contatto visivo con il nostro interlocutore è grazie a tali informazioni trasmesse attraverso il canale vocale che, oltre a decodificare il messaggio verbale, siamo anche in grado di inferire e determinare tutta una serie di informazioni legate al nostro interlocutore (tra cui età, sesso, provenienza ecc.). Allo stesso modo, attraverso la voce, comunichiamo le nostre intenzioni emotive: durante una discussione con un altro individuo, siamo in grado di capire, condividendo o meno tutta una serie di informazioni, se il nostro interlocutore è triste, arrabbiato o altro, semplicemente facendo riferimento a “come” egli dice determinate cose. Si rammenta a tal proposito quanto affermato da Sawamura *et al.* (2007: 2113):

“During daily conversation, we can perceive emotions via speech even if we cannot understand the linguistic meaning.”

Che gli esseri umani siano in grado di decodificare correttamente produzioni emotive con gradi di successo che vanno ben oltre la pura casualità è un dato oramai ben consolidato che emerge da numerosi studi sull’argomento anche nel caso in cui si affronti la tematica del parlato emotivo in lingue diverse dalla propria, una capacità diffusamente documentata dalle ricerche condotte sino ad oggi in contesti cross-linguistici e cross-culturali di cui si è più sopra fornita un’ampia rassegna (vedi capitolo 7)

Stando alla già richiamata meta-analisi condotta da Laukka (2004) emerge, inoltre, come il *decoding* cross-culturale sia inferiore al *decoding* intra-culturale del 7% facendo rilevare la presenza di quello che viene definito *in-group advantage* nel riconoscimento di emozioni nella propria lingua.

Il dato riportato da Laukka (*op. cit.*) confermerebbe anche l’idea di Scherer, Banse, Wallbott (2001: 78) secondo i quali,

“it seems reasonable to assume that the recognition of vocal emotion expressions might work across language and culture boundaries”.

Un altro dato a cui si è spesso fatto riferimento, evidenzia come l’aspetto del *decoding* cross-linguistico e cross-culturale sia stato investigato in varie forme e modi: vi sono ricerche che hanno preso in esame contesti culturali nettamente contrapposti e spesso agli antipodi (americani vs. giapponesi, o ancora vs. cinesi); utilizzando stimoli vocali composti da frasi *meaningless, non-sense*, semplici parole ecc., filtrati e non, e che, in ciascun esperimento non sono stati raccolti in più di tre lingue. Come rilevato, pochissimi tra questi, infine, si sono occupati di sole lingue europee e ancora di meno sono quegli studi che hanno preso in considerazione stimoli in lingua italiana o ascoltatori italiani.

È per questa ragione che si è ritenuta di fondamentale importanza la conduzione di un esperimento di *decoding* di produzioni vocali emotive in quattro lingue europee. In questa sezione si intende, perciò, verificare l'abilità da parte di soggetti di madrelingua italiana nel decodificare emozioni vocali "posate" prodotte da soggetti *attori* e *naif* in quattro lingue europee (italiano, francese, inglese e tedesco).²⁶⁹

Gli obiettivi principali alla base della presente proposta saranno pertanto quelli di:

- a. somministrare degli esperimenti di tipo percettivo a soggetti di madrelingua italiana per appurare la loro capacità di decodificare espressioni vocali emotive nella propria lingua e in lingue differenti dalla propria (inglese, francese e tedesco);
- b. stabilire in che misura il grado di conoscenza e competenza della lingua presentata possa influire sul riconoscimento delle produzioni emotive presentate verificando, al contempo, quanto ipotizzato da Scherer, Banse, Wallbott (2001: 88);²⁷⁰
- c. verificare se vi siano differenze di riconoscimento riconducibili al sesso dei soggetti ascoltatori (*gender-effects*): in alcuni studi è stata, ad esempio, documentata la maggiore abilità da parte di soggetti di sesso femminile nel decodificare emozioni espresse attraverso il volto, anche se per il canale vocale questo dato non risulta particolarmente significativo;²⁷¹
- d. verificare se vi sia effettivamente un *in-group advantage* nel riconoscimento delle emozioni vocali come evidenziato in tutti gli studi più sopra richiamati;
- e. verificare se vi siano emozioni maggiormente riconosciute secondo quanto già rilevato in fase di validazione delle registrazioni parte delle quali saranno utilizzate in questa fase dell'indagine;
- f. verificare se vi siano differenze di riconoscimento tra produzioni realizzate da *attori* o da soggetti *naif*: da quanto emerso da alcuni studi, ad es. dallo studio

²⁶⁹ È comunque in programma la continuazione dello studio presentato in questa sezione con la replica dell'esperimento di *decoding* utilizzando anche soggetti ascoltatori di madrelingua francese, inglese e tedesca. Ciò consentirà di avere una comprensione complessiva del problema affrontato soddisfacendo anche la proposta di Scherer, Banse, Wallbott (2001), pag. 88, che auspicano come "[...] *encoders and decoders from several different countries would need to be studied, allowing the construction of an encoder-decoder-emotion matrix and to test whether decoders from the countries involved would recognize emotion portrayals by encoders from their own countries most accurately*".

²⁷⁰ Ovvero: "*It seems that as soon as vocal expressions, other than pure nonlinguistic affect bursts, are used, segmental and suprasegmental aspects of language affect encoding and decoding of emotion*".

²⁷¹ Per questo ed altri aspetti si veda nello specifico il capitolo 8 "Gender differences in emotion Processes" in Niedenthal, P. M., Krauth-Gruber, S., Ric, F. (2006), *The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches*, Principles of Social Psychology series, New York: Psychology Press, pp. 271-304.

di Wilting, Kraemer, Swerts (2006), in cui è emersa una maggiore riconoscibilità delle emozioni prodotte da soggetti attori, soprattutto per emozioni negative;

- g. dalla somministrazione di una serie di domande ai soggetti ascoltatori si intende inoltre verificare se alcune impressioni personali sullo svolgimento del test trovano effettivamente riscontro nei valori di riconoscimento relativi agli stimoli somministrati.

Lo sviluppo di questa fase della ricerca avverrà in contemporanea con il terzo *step*.

8.3 Step 3: analisi dei correlati acustici delle produzioni emotive mistilingui

Se, come tra l'altro emerso negli studi sin qui richiamati, vi è nell'uomo una certa capacità nel riconoscere emozioni trasmesse attraverso la sola voce anche in contesto cross-culturale a livello cross-linguistico, e se tale dato dovesse venire confermato, come ci si aspetta che sia, anche nel presente caso, allora è assai evidente che l'informazione acustica abbia la sua importanza nell'influenzare o nell'indirizzare l'ascoltatore nel riconoscimento di meri stimoli vocali emotivi anche in presenza di lingue diverse dalla propria.

In aggiunta a ciò, per verificare se eventuali differenze di riconoscimento da parte del gruppo di ascoltatori siano correlate alle diverse caratterizzazioni acustiche delle singole produzioni emotive presentate, in questa fase sarà operata una analisi acustica delle produzioni presentate ai soggetti ascoltatori nel secondo *step*.

In tale fase si procederà pertanto all'estrazione di una serie di indici acustici nel dominio della frequenza, dell'intensità e del tempo allo scopo di:

1. caratterizzare e definire quelli che sono i profili acustici delle emozioni codificate in ciascuno degli stimoli in ciascuna delle lingue esaminate;
2. verificare se le predizioni formulate da Scherer (1984, 1986, 2003; Banse, Scherer, 1996) trovano supporto anche nella presente indagine;
3. valutare se negli stimoli presentati agli ascoltatori e analizzati in questa fase vi siano differenze a livello acustico riconducibili alle eventuali diverse percentuali di riconoscimento;
4. stabilire se vi siano sostanziali differenze tra soggetti *attori* e soggetti *naïf* nella codifica o produzione di emozioni posate allo scopo di verificare l'assunto più volte richiamato secondo cui le espressioni emotive prodotte da soggetti *attori* sarebbero da ritenersi più intense o più marcatamente caratterizzate rispetto a quelle prodotte da soggetti *naïf*.

9 Step 1: Progettazione, raccolta e creazione di un corpus emotivo mistilingue

In questo capitolo, e nei paragrafi che seguono, verranno illustrate le scelte di tipo metodologico che hanno portato alla progettazione, alla raccolta e alla creazione del corpus di parlato emotivo mistilingue per le quattro lingue europee, operazione propedeutica e necessaria all'attuazione delle successive fasi della ricerca qui intrapresa.²⁷²

9.1 La frase standard

Avendo in mente un confronto di produzioni vocali emotive in contesto cross-linguistico si è optato per una frase standard, anche definita *carrier sentence*²⁷³, che consentisse una adeguata confrontabilità dei dati riducendo in *primis* la variabilità inter-parlatore nella produzione delle emozioni vocali. Al fine di garantire un certo livello di confrontabilità dei dati (sia a livello acustico, sia a livello percettivo) sono stati fissati *a-priori* dei *desiderata*, secondo cui la frase scelta doveva essere:

- semanticamente neutra se estrapolata da uno specifico contesto. Nonostante gli sforzi non è stato possibile soddisfare questo requisito per il semplice fatto che qualsiasi frase, generata e trasmessa durante un atto comunicativo, porta con sé una serie di informazioni che non è possibile escludere *a-priori*. Il semanticamente neutro è stato perciò ricondotto alla possibilità di trovare una frase che potesse essere inserita e presentata in qualsiasi contesto emotivo in modo coerente con l'emozione intesa;
- coerente con le situazioni rappresentate per ciascuna emozione. La frase scelta doveva essere in linea con il costrutto della scena o della situazione presentata per

²⁷² I risultati di cui al presente capitolo rappresentano la versione estesa della pubblicazione a nome Galatà, V., Romito, L. (in stampa), Un corpus sperimentale per lo studio cross-linguistico europeo delle emozioni vocali, in: *Atti del 5° Convegno AISV - "La dimensione Temporale del parlato"*, Università di Zurigo, Svizzera, 4-6 febbraio 2009.

²⁷³ O ancora *control cluster* come la definiscono Williams, Stevens (1972).

ciascuna emozione, affinché la stessa non ingenerasse nel soggetto situazioni di dubbio o diffidenza rispetto al materiale da produrre;

- di utilizzo comune in ciascuna lingua esaminata. La frase doveva rispettare quelle che sono rappresentate dalle consuetudini terminologiche e di costruito di ciascuna lingua che la rendesse accettabile agli *encoder* prima ed ai *decoder* dopo;
- “facile” da analizzare. Con questa terminologia si fa riferimento alla possibilità di individuare con un certo grado di facilità e sicurezza quelli che sono i confini frasali, nello specifico l’inizio della frase che sarebbe dovuto avvenire con un fono possibilmente sonoro. Tale requisito è motivato dalla possibilità di stabilire con esattezza, nella fase di segmentazione ed etichettatura, l’inizio della produzione all’interno del segnale registrato;²⁷⁴
- contenere una pausa. Si è ritenuto in questo caso significativa la presenza e l’inserimento di una pausa all’interno della frase, al fine di valutare, nelle successive fasi, se nelle quattro lingue esaminate vi sia un diverso utilizzo o gestione dei tempi legati alle pause e ai silenzi.

Date queste premesse, e sulla base di precedenti esperienze di ricerca e scelte operate da altri studiosi²⁷⁵, con l’aiuto di docenti di madrelingua²⁷⁶ si è passati all’individuazione di una frase standard nelle quattro lingue esaminate²⁷⁷ giungendo alla seguente scelta:

-  Non è possibile. Non ci posso credere.
-  Oh là là. C’est incroyable.
-  It can’t be. I cannot believe it.
-  Das ist nicht möglich. Ich kann es nicht glauben.

²⁷⁴ Cosa che sarebbe sostanzialmente impossibile con foni di tipo oclusivo che sono caratterizzati per loro natura dalla presenza di una fase di occlusione, o di silenzio, non rilevabile in posizione di inizio di frase.

²⁷⁵ Si fa qui riferimento a Walbott, Scherer (1986) che scelsero la frase “*Ich kann es nicht glauben.*” e ad Anolli, Ciceri (1992) (ma anche Anolli *et al.* 2008a, 2008b) che scelsero la frase “*Non è possibile, non ora.*”.

²⁷⁶ Tutti con esperienza nel campo dell’insegnamento e della traduzione e in servizio presso la Facoltà di Lettere e Filosofia dell’Università della Calabria.

²⁷⁷ Scelta operata anche al fine di poter valutare il peso della competenza della lingua in un esperimento percettivo di tipo cross-linguistico sulle emozioni vocali.

9.2 I testi emotigeni o scenari

Per l'elicitatione delle emozioni desiderate si è fatto riferimento al paradigma degli scenari. Gli scenari sono rappresentati sotto forma di testi narrativi, per la stesura dei quali si è fatto ricorso ad alcuni esempi presenti in letteratura. Nello specifico sono stati utilizzati, adattandoli in alcune parti, i testi emotigeni creati e riportati in Anolli, Ciceri (1992) e successivamente utilizzati anche in Anolli *et al.* (2008a, 2008b). L'adattamento dei testi si è reso necessario al fine di consentirne una più agevole traduzione nelle lingue oggetto di studio.²⁷⁸

Sono stati utilizzati complessivamente 6 scenari con riferimento alle emozioni definite da Ekman (1992) come *basic* (*happiness, anger, fear, sadness, disgust, surprise*). Poiché negli studi a cui si è fatto riferimento per gli scenari, l'emozione di sorpresa non era stata trattata, si è reso necessario procedere alla stesura *ex novo* del relativo testo seguendo le linee guida riportate nel § 6.2.4. L'utilizzo degli scenari ha, inoltre, consentito di ridurre al minimo il ricorso ad etichette verbali di tipo emozionale (rabbia, tristezza ecc.) nelle quattro lingue, evitando in tal modo le problematiche a cui si è accennato al § 6.5. Tutti i testi creati ed utilizzati nella presente indagine sono riportati in Appendice.

9.3 Raccolta del corpus: modalità e caratteristiche

La raccolta del corpus è stata effettuata in ambiente insonorizzato con un microfono direzionale mod. *Sennheiser e835* direttamente in formato *.wav con l'ausilio di un registratore digitale mod. *Edirol R-09* a 44.1kHz 16-bit mono.²⁷⁹

Il microfono, sorretto da un'asta microfonica, è stato posizionato ad una distanza di ca. 10 cm²⁸⁰ e con un'angolatura di 45° rispetto alla fonte di emissione, come indicato in Figura 36 di seguito riportata a titolo esemplificativo.

²⁷⁸ Particolare attenzione è stata data in questo caso all'utilizzo di termini di uso comune, evitando volutamente un linguaggio "forbito" che avrebbe avuto ripercussioni sulla caratteristica di spontaneità in fase di produzione da parte dei soggetti.

²⁷⁹ Le registrazioni sono state effettuate in parte nella camera insonorizzata (marca *Amplifon 2x2*) in dotazione al Laboratorio di Fonetica dell'Università della Calabria (<http://www.linguistica.unical.it/labfon/Home.htm>), in parte nella camera insonorizzata del KTH di Stoccolma durante un soggiorno presso il *Centre for Speech Technology* (CTT) in qualità *guest researcher* (<http://www.speech.kth.se/ctt/>) e in parte presso lo studio di registrazione internazionale *Studio Colosseo* con sede in Roma (<http://www.studiocolosseo.com/>).

²⁸⁰ Sebbene il microfono sia stato preventivamente posizionato alla stessa distanza per tutti i soggetti, mantenendo costante la distanza prefissata, è stato impossibile impedire ai soggetti di muoversi, causando un'oscillazione della distanza dal microfono di ± 3 cm.



Figura 36: Disposizione del microfono con angolatura di 45° rispetto alla fonte di emissione

La procedura di raccolta delle produzioni di parlato emotivo è stata suddivisa in quattro fasi, una consecutiva all'altra.²⁸¹ Ciascuna delle quattro fasi risulta caratterizzata da una specifica modalità di acquisizione del materiale scaturita dalla somministrazione al soggetto di particolari istruzioni o informazioni.

Tutte le scelte, le strategie e le impostazioni metodologiche sono state inizialmente verificate e testate con l'aiuto di due attori italiani. Dopo una prima sommaria analisi del materiale acquisito, si sono rivelati necessari piccoli aggiustamenti nell'impostazione della procedura di raccolta. Di seguito le caratteristiche di ciascuna delle modalità di raccolta del corpus in questione nella loro versione definitiva:

- Modalità A: al soggetto è stato chiesto di leggere i 6 brani (uno per ciascuna emozione studiata) in modo spontaneo e naturale senza ulteriori informazioni;
- Modalità B: dopo aver esplicitato al soggetto il motivo della registrazione e dopo un *feedback* da parte del soggetto riguardo l'emozione a cui si faceva riferimento in ciascun testo²⁸², al soggetto è stato chiesto di rileggere il testo, sempre in modo spontaneo e naturale, rendendosi partecipe della situazione descritta (soprattutto nelle parti in cui era prevista la forma dialogica);
- Modalità C: al soggetto è stato chiesto di produrre la frase standard per ciascuna delle emozioni per almeno 4 volte²⁸³. Al soggetto è stato, inoltre, esplicitamente chiesto di fare riferimento alle emozioni descritte nei testi precedentemente letti: in questa fase è stato messo in atto quello che potrebbe essere definito uno "pseudo-metodo Stanislavskij", sostituendo alla vita e all'esperienza personale

²⁸¹ L'intera procedura, con l'attuazione di tutte e quattro le modalità di raccolta, ha richiesto tra i 40 e i 60 minuti a soggetto.

²⁸² Tale scelta è stata dettata dalla necessità di validare anche i brani adattati da Anolli, Ciceri (1992). È stato, infatti, in questa fase che è stato possibile riscontrare la presenza per il tedesco di due termini per quella che in italiano viene definita "tristezza". Nel caso specifico, con riferimento allo scenario proposto, quella che era stata erroneamente indicata nel tedesco come *Traurigkeit* era invece *Trauer*.

²⁸³ Al soggetto è stato chiesto di fare una breve pausa tra una produzione e l'altra al fine di evitare influenze di tipo stilistico causate dalla concatenazione delle frasi prodotte.

richiesta dal metodo per l'autoinduzione dello stato emotivo, la descrizione contenuta nei brani presentati per ciascuna delle emozioni investigate.²⁸⁴ A fine sessione è stata data facoltà al soggetto di ripetere l'intera procedura relativa alla presente modalità di raccolta qualora non fosse stato soddisfatto dalla propria *performance*;²⁸⁵

- Modalità Neutra: a conclusione delle tre fasi di raccolta delle produzioni verbali propriamente emotive, si è proceduto alla raccolta delle produzioni neutre per la stessa frase. Al soggetto è stato chiesto di leggere, per almeno quattro volte, ciascuna delle 5 frasi riportate su singoli fogli.²⁸⁶ L'istruzione data al soggetto è stata quella di leggere le frasi nel modo più possibile neutro e senza alcuna caratterizzazione emotiva: l'esempio dato al soggetto è stato quello di leggere le frasi come se si trattasse di una lista di ingredienti per una ricetta culinaria.

In alcuni casi, a fine sessione di registrazione di ciascuna fase (prevalentemente per la modalità A e B), si è reso necessario far ripetere al soggetto qualcuno dei paragrafi contenente la frase standard, in quanto, nella lettura, erano stati inseriti nella frase standard suoni diversi da quelli previsti (in alcuni casi anche diverse parole).²⁸⁷

L'unica informazione fornita ai soggetti in fase di reclutamento riguardava esclusivamente l'oggetto di studio cross-linguistico delle emozioni vocali nell'ambito di una tesi di dottorato. Qualsiasi altra richiesta di informazioni in merito allo svolgimento dell'esperimento è stata declinata a tutela del *setting* sperimentale, spiegando che qualsiasi altra informazione aggiuntiva avrebbe vanificato il costrutto sperimentale della ricerca.²⁸⁸

Per ciascuna lingua sono state acquisite registrazioni per le due categorie di soggetti, rispettivamente *naif* e *attori* professionisti, tutti di sesso maschile. La tabella Tabella 15 di seguito riporta un quadro relativo al numero e l'età media (tra parentesi la deviazione

²⁸⁴ È stato così possibile ridurre al minimo la variabilità legata alla situazione elicitante che in altri casi sarebbe stata inevitabilmente legata al vissuto personale dei singoli come prevede appunto il metodo "Stanislavskij".

²⁸⁵ Durante le prime sessioni di registrazione è stato possibile evidenziare come il soggetto si rendesse effettivamente conto delle produzioni prodotte in questa fase solo dopo aver esperito, attraverso la produzione verbale, le emozioni richieste, manifestando la volontà di ripetere alcune delle produzioni emotive realizzate e per le quali era stato espresso un certo grado di insoddisfazione.

²⁸⁶ La frase di interesse è stata inserita all'interno dei 5 fogli e quindi delle 5 frasi. Anche in questo caso al soggetto è stato chiesto di fare una breve pausa tra una produzione e l'altra al fine di evitare influenze di tipo stilistico causate dalla concatenazione delle frasi prodotte.

²⁸⁷ Per la caratteristica delle modalità di raccolta delle produzioni nella modalità A e B, sebbene si tratti di parlato letto, le stesse possono essere considerate come produzioni emotive spontanee o semi-spontanee dal momento che i soggetti coinvolti ignoravano quale fosse la frase oggetto di studio e quale fosse, di fatto, l'obiettivo specifico delle due modalità. Di contro, le produzioni raccolte nella modalità C possono essere considerate a pieno titolo come produzioni emotive posate o simulate, in quanto in tale modalità l'attenzione del soggetto è stata fatta convergere esclusivamente sulla frase proposta.

²⁸⁸ Ai soggetti coinvolti è stato comunque spiegato che a fine sessione le finalità del lavoro si sarebbero palesate e che si sarebbe, infine, data risposta a tutte le loro curiosità o domande.

standard) dei soggetti *naif* e degli *attori* coinvolti in questa fase per l'acquisizione delle produzioni emotive in ciascuna lingua:

Lingua	Soggetti	Numero	Eta' media (SD)
Italiano (it)	attori	5	35,8 (5,2)
	naif	6	25,8 (1,6)
Francese (fr)	attori	3	43,7 (11,0)
	naif	6	29,5 (12,8)
Inglese (en)	attori	2	46,0 (8,5)
	naif	6	45,0 (10,9)
Tedesco (de)	attori	2	50,5 (2,1)
	naif	9	26,0 (3,9)

Tabella 15: Riepilogo informazioni sul campione dei soggetti registrati suddivisi per lingua e per tipologia

Prima e durante le sessioni di registrazione, tutte le istruzioni sono state fornite a ciascun soggetto nella propria lingua madre in forma sia scritta che orale.

La procedura di raccolta appena descritta è stata attuata per ciascun soggetto di ciascuna lingua per arrivare, a conclusione delle operazioni di etichettatura (*cfr.* § 9.4 a seguire), ad un totale di 40 produzioni a soggetto così suddivise:

- Modalità A: 6 produzioni emotive (1 per emozione);
- Modalità B: 6 produzioni emotive (1 per emozione);
- Modalità C: 24 produzioni emotive (4 per ciascuna emozione);
- Modalità Neutra: 4 produzioni neutre.

Di seguito, alla fase di raccolta delle produzioni emotive per ciascun soggetto, si è provveduto ad isolare ed etichettare le singole produzioni contenute nel *continuum* delle registrazioni acquisite.

9.4 Etichettatura del corpus

L'etichettatura del corpus, intesa qui nell'accezione di isolamento mediante segmentazione della frase standard raccolta per ciascun soggetto nelle quattro modalità, è stata effettuata manualmente con ascolto in cuffia e con l'ausilio di un software *open source* denominato Praat²⁸⁹. Particolare attenzione è stata prestata, in questa fase, all'individuazione

²⁸⁹ *Cfr.* Boersma, P., Weenink, D. (2009), *Praat: doing phonetics by computer* [Computer program], retrieved from <http://www.praat.org/>. Il programma, oltre a consentire di visualizzare in contemporanea oscillogramma e sonogramma del segnale analizzato rendendo più agevole e precisa la segmentazione e l'etichettatura, permette anche di non intaccare e/o modificare il file originale, riportando tutte le operazioni effettuate sull'originale in un file di testo denominato *Textgrid*.

dei punti di attacco e fine frase, avendo cura di operare la segmentazione della porzione con il taglio del segnale sonoro in corrispondenza dello *zero-crossing* (vedi Figura 37).

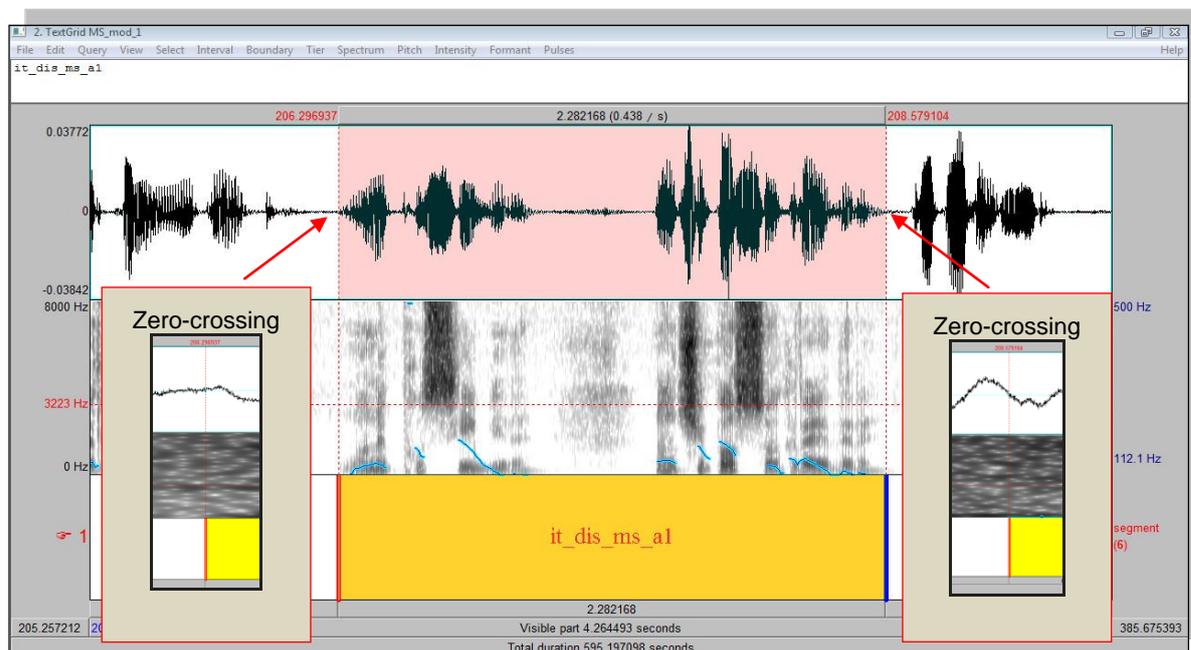


Figura 37: Esempio di schermata relativa alla segmentazione e all'etichettatura delle produzioni

L'etichettatura della porzione di segnale di interesse fornisce il nome del file assegnato alla porzione nella fase di esportazione della stessa in un singolo file. L'etichettatura adottata in fase di segmentazione codifica tutte le informazioni disponibili per quella data porzione, informazioni standardizzate secondo uno schema a sei campi in termini di [lingua]_[emozione]_[categoria]_[iniziali]_[(modalità)occorrenza].wav, dove:

1. [lingua] = campo di due lettere che corrisponde alla lingua codificata:

- it (italiano);
- fr (francese);
- en (inglese);
- de (tedesco);

2. [emozione] = campo di tre lettere che corrisponde all'emozione codificata:

- ang (collera calda);
- dis (disgusto);
- fea (paura);
- joy (gioia);
- neu (neutro);
- sad (tristezza);
- sur (sorpresa);

3. [categoria] = campo di una lettera che corrisponde alla categoria di appartenenza del soggetto utilizzato per l'*encoding* dell'emozione :
 - a (attore);
 - n (naif);
4. [iniziali] = campo di due lettere corrispondente alle iniziali del soggetto registrato;
5. (modalità) = campo di una lettera che corrisponde alla modalità di raccolta della produzione emotiva codificata:
 - a (lettura del brano in Modalità A);
 - b (lettura del brano in Modalità B);
 - c (lettura della sola frase in Modalità C);
 - campo non contemplato nel caso di produzione neutra in quanto raccolto in un'unica e sola modalità (Modalità Neutra);
6. [occorrenza] = campo di una cifra relativa all'occorrenza della frase prodotta.

A titolo esemplificativo vengono di seguito forniti alcuni esempi relativi alla predetta etichettatura dei *files*: *it_neu_n_gb_1.wav*; *fr_ang_a_hd_c3.wav* e così via.

A conclusione della fase di segmentazione ed etichettatura dell'intero corpus sono risultate nell'insieme 1560 produzioni come di seguito meglio dettagliato (vedi Tabella 16):

	Soggetti naif	Attori	Totale	N° produzioni
Italiano (it)	6	5	11	440
Francese (fr)	6	3	9	360
Inglese (en)	6	2	8	320
Tedesco (de)	9	2	11	440
<i>Totale</i>	<i>27</i>	<i>12</i>	<i>39</i>	<i>1560</i>

Tabella 16: Riepilogo produzioni complessivamente raccolte

9.5 Procedura di validazione del corpus

Come sottolinea Magno Caldognetto (2002: 212) la validazione delle registrazioni acquisite devono necessariamente essere validate attraverso dei test percettivi e di adeguatezza al fine di:

“selezionare le produzioni riconosciute esattamente e con il punteggio più elevato da parte degli ascoltatori [...]”

Ciò si rende necessario non solo per assicurare la significatività di tutte quelle operazioni legate al prosieguo della ricerca e della sperimentazione intrapresa (misurazioni acustiche, prove percettive ecc.), ma anche per assicurare e garantire un certo grado di

significatività e confrontabilità dei risultati con quelli conseguiti in altre ricerche. D'altro canto, tale operazione si rende assolutamente necessaria nel caso si presentino produzioni emotive di una lingua a membri appartenenti ad un'altra lingua o cultura.

Anche nel caso della validazione allo stato attuale non esistono, per quanto sia noto, procedure standardizzate. Per questa ragione, per la validazione delle registrazioni acquisite per ciascuna lingua, sono stati somministrati due test di adeguatezza:

- uno relativo all'identificazione dell'emozione espressa nello stimolo presentato (nel prosieguo identificato per comodità con la sigla "T1") con una scelta da effettuare sulla base di un *set* prefissato di etichette emozionali (tristezza, collera, paura, neutro, gioia, sorpresa e disgusto) con in aggiunta la possibilità di esprimere il grado di certezza della risposta data (incerto, certo)²⁹⁰. Scopo di questo test è stato quello di escludere tutte quelle produzioni che potevano suscitare confusione e che potevano essere considerate emotivamente ambigue;
- uno relativo alla rappresentatività dello stimolo presentato (nel prosieguo identificato con la sigla "T2") contestualmente all'indicazione dell'emozione intesa in fase di *encoding* (rappresentatività espressa su una scala di giudizio di tipo pari a quattro gradi).²⁹¹

La scelta di somministrare un secondo test di validazione, come quello sulla rappresentatività di un determinato stimolo in funzione dell'emozione intesa, è riconducibile a due motivazioni:

- la prima è sostanzialmente dettata dall'aver in mente una ulteriore selezione degli stimoli da utilizzare nell'esperimento percettivo di tipo cross-linguistico (e per il quale tutto il lavoro qui presentato si è reso necessario): a parità di identificazione corretta di una o più produzioni per una stessa emozione espressa da un soggetto, sarebbe stato così possibile scegliere quella giudicata più rappresentativa da parte dei giudici ascoltatori;
- la seconda è dovuta alla verifica di una ipotesi, ovvero che uno stimolo non correttamente identificato difficilmente avrebbe avuto una valutazione alta in termini di rappresentatività.

²⁹⁰ Trattandosi di un test di tipo *forced-choice* è stata data agli ascoltatori la possibilità di esprimere il grado di confidenza (certezza/incertezza) sulla risposta data per evitare che gli stessi identificassero con l'opzione "neutro" la condizione di incertezza o di dubbio.

²⁹¹ È stata qui adottata una scala di giudizio *pari* per obbligare il soggetto a prendere una posizione chiara ed univoca evitando situazioni di dubbio o di neutralità.

Per una validazione omogenea dell'intero *corpus* e per evitare di somministrare troppi stimoli ad un unico soggetto,²⁹² e poiché per ciascuna lingua sono stati utilizzati gli stessi soggetti che hanno partecipato alla fase di acquisizione, a ciascun soggetto sono stati somministrati i due test includendo le registrazioni prodotte dal soggetto ascoltatore medesimo e quelle di altri due soggetti per la stessa lingua. I test contenenti gli stimoli da far ascoltare ai soggetti sono stati generati sulla base di una matrice come quella presentata in Tabella 17.

DECODER		A ¹	B ¹	C ¹	D ¹	E ¹
ENCODER	A					
	B					
	C					
	D					
	E					

Tabella 17: Modalità di generazione dei set di ascolto utilizzati per i due test di validazione (ordine di lettura dall'alto verso il basso)

Si supponga, per ciascuna lingua, la presenza di un set di voci A, B, C, D, E come riportato in Tabella 17, dove la colonna di sinistra si riferisce alle produzioni verbali emotive, mentre la riga in alto si riferisce all'ascoltatore (che nel presente caso, come nei test successivamente condotti, coincidono con gli *encoder*). Sulla base della matrice riportata in Tabella 17, le produzioni del soggetto A, oltre ad essere ascoltate da se stesso (A¹) vengono ascoltate dai soggetti B¹ ed E¹; le produzioni del soggetto B, oltre ad essere ascoltate da se stesso (B¹) vengono ascoltate dai soggetti A¹ e C¹; le produzioni del soggetto C, oltre ad essere ascoltate da se stesso (C¹) vengono ascoltate dai soggetti B¹ e D¹; le produzioni del soggetto D, oltre ad essere ascoltate da se stesso (D¹) vengono ascoltate dai soggetti C¹ ed E¹; infine, le produzioni del soggetto E, oltre ad essere ascoltate da se stesso (E¹) vengono ascoltate dai soggetti D¹ e A¹ chiudendo in questo modo la batteria dei test. In questo modo al soggetto A¹ verranno somministrate le produzioni sue (quindi A) nonché quelle di B ed E e così via. Secondo tale costrutto, ciascuna produzione di ciascun soggetto viene valutata complessivamente da tre soggetti.

I due test sono stati quindi implementati con la funzione “ExperimentMFC” del software Praat e somministrati in cuffia: gli stimoli sono stati opportunamente randomizzati con una apposita *routine* all'interno dello stesso programma.²⁹³ In entrambi i casi il test era preceduto da una schermata con poche e semplici istruzioni sul compito da svolgere. Anche in

²⁹² Si consideri che per ciascun soggetto sono stati complessivamente acquisiti 40 stimoli.

²⁹³ Per evitare il calo di attenzione da parte dei giudici ascoltatori, in entrambi i test, è stata prevista una pausa ogni 30 stimoli presentati.

questo caso il tutto è stato svolto nella lingua madre dei soggetti ascoltatori (vedi Figura 38 e Figura 39).²⁹⁴

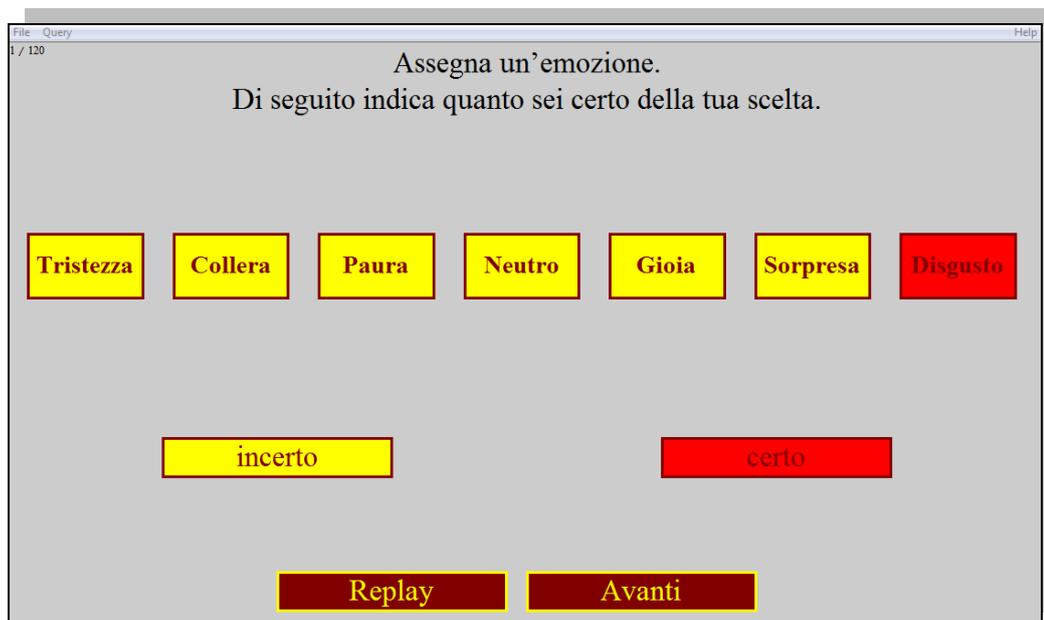


Figura 38: Schermata relativa al test di validazione T1 sull'identificazione dell'emozione presentata

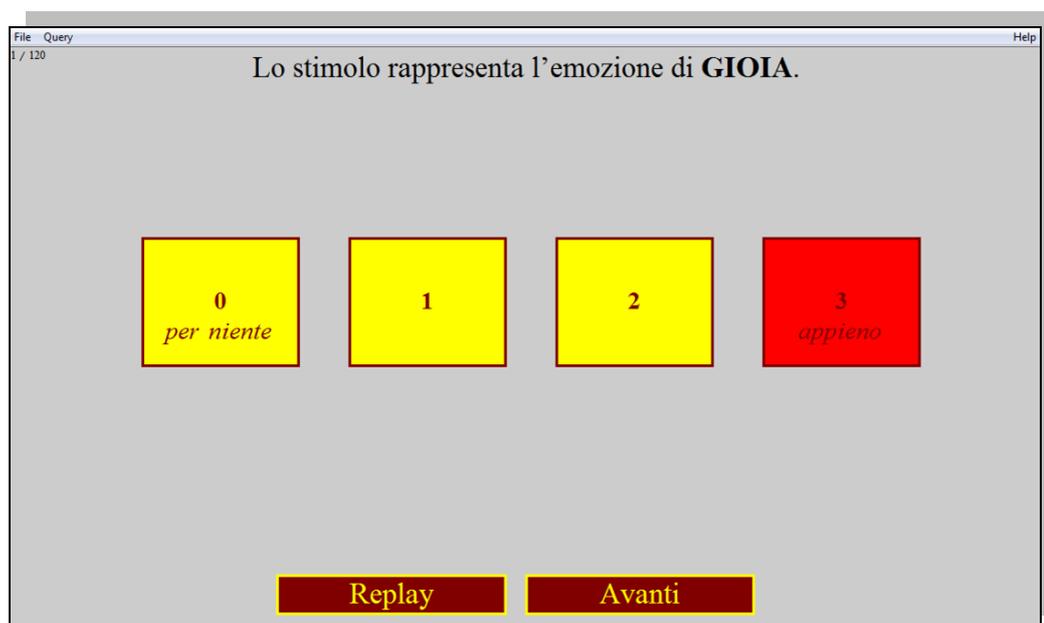


Figura 39: Schermata relativa al test di validazione T2 sulla rappresentatività dell'emozione presentata

²⁹⁴ In Appendice VII, vengono forniti, rispettivamente, gli script compilati e somministrati per mezzo del software Praat per le schermate relative ai test di validazione T1 in Figura 38 e T2 in Figura 39 per la lingua italiana. Analoga operazione si è resa necessaria per le lingue straniere, dove alle istruzioni in italiano sono state sostituite le istruzioni in lingua straniera.

9.6 Risultati della validazione del corpus

Vengono di seguito riportati i risultati relativi ai due test di validazione adottati per verificare la bontà e l'adeguatezza delle registrazioni di parlato emotivo raccolte per ciascuna delle quattro lingue secondo le modalità sopra riportate, registrazioni che costituiranno il corpus di parlato emotivo mistilingue qui presentato, parte del quale sarà utilizzato per l'attuazione degli altri due *step* della ricerca.

Prima di intraprendere qualsiasi analisi ci si è dapprima chiesti se e in che misura i risultati dei due test di validazione, rispettivamente T1 e T2, fossero tra loro correlati. Poiché era stato ipotizzato che uno stimolo non correttamente identificato nel T1 difficilmente avrebbe avuto una valutazione alta in termini di rappresentatività nel T2, per verificare la validità di tale ipotesi si è proceduto al calcolo di un coefficiente di correlazione sui valori medi delle risposte fornite per ciascuno stimolo nei due test di validazione: per tutte e quattro le lingue è stata rilevata una correlazione superiore a 0.5 con un livello di significatività $p = 0.01$ supportando quella che era l'ipotesi di partenza con valori rispettivamente di $r = 0.672$ per l'italiano, $r = 0.706$ per il francese, $r = 0.532$ per l'inglese e $r = 0.705$ per il tedesco.

9.6.1 Test di validazione relativo all'identificazione delle emozioni (T1)

Il primo dato ad emergere dai test di validazione relativi all'identificazione delle emozioni presentate ai giudici ascoltatori è sicuramente legato alla differenza dei valori di corretta identificazione delle emozioni espresse in base alla modalità di raccolta delle produzioni verbali emotive. Infatti, come si evince dalle matrici di confusione generate sulla base del test di identificazione T1, per ciascuna delle quattro lingue esaminate e per ciascuna delle modalità di raccolta delle produzioni (vedi Tabelle 18-21), emerge sostanzialmente come vi sia, da parte dei giudici ascoltatori, un certo grado di abilità nell'identificare produzioni emotive nella propria lingua con percentuali nettamente al di sopra della pura casualità²⁹⁵. Indifferentemente dalle modalità, si registrano valori medi di corretta identificazione del 56% per l'italiano, del 51% per il francese, del 44% per l'inglese e del 49% per il tedesco. Si tratta di percentuali indubbiamente alte se si considera il fatto che ai giudici ascoltatori è stato presentato, in questo caso, materiale in forma "grezza".

²⁹⁵ Pari a ca. il 14,3% in un test con sette opzioni.

ITALIANO																	
Valori complessivi		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta A		sad	ang	fea	neu	joy	sur	dis
	sad	74%	2%	4%	14%	1%	2%	5%		sad	52%	3%	6%	27%	0%	3%	9%
	ang	2%	63%	3%	14%	1%	7%	13%		ang	6%	15%	6%	42%	0%	9%	21%
	fea	12%	2%	44%	15%	4%	13%	11%		fea	3%	0%	15%	42%	6%	21%	12%
	neu	6%	1%	0%	90%	2%	0%	2%		neu	6%	1%	0%	90%	2%	0%	2%
	joy	4%	3%	8%	8%	40%	35%	3%		joy	9%	0%	6%	24%	21%	33%	6%
	sur	9%	4%	6%	16%	5%	40%	20%		sur	6%	0%	9%	30%	0%	36%	18%
	dis	6%	13%	5%	13%	3%	21%	39%		dis	3%	15%	3%	33%	0%	33%	12%
Modalità di raccolta B		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta C		sad	ang	fea	neu	joy	sur	dis
	sad	76%	0%	9%	15%	0%	0%	0%		sad	80%	2%	2%	10%	1%	2%	5%
	ang	3%	45%	3%	18%	3%	9%	18%		ang	0%	79%	2%	5%	0%	5%	9%
	fea	6%	3%	42%	15%	3%	12%	18%		fea	16%	2%	52%	8%	4%	11%	8%
	neu	6%	1%	0%	90%	2%	0%	2%		neu	6%	1%	0%	90%	2%	0%	2%
	joy	3%	3%	9%	15%	42%	24%	3%		joy	2%	3%	8%	2%	45%	38%	2%
	sur	15%	6%	6%	18%	0%	27%	27%		sur	8%	5%	5%	12%	8%	45%	18%
	dis	0%	15%	6%	18%	0%	18%	42%		dis	6%	9%	0%	9%	6%	18%	52%

Tabella 18: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue italiano²⁹⁶

FRANCESE																	
Valori complessivi		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta A		sad	ang	fea	neu	joy	sur	dis
	sad	67%	3%	6%	14%	2%	6%	3%		sad	44%	7%	7%	33%	0%	4%	4%
	ang	1%	56%	5%	11%	5%	17%	7%		ang	4%	22%	11%	30%	0%	15%	19%
	fea	22%	2%	38%	11%	3%	22%	1%		fea	0%	0%	44%	30%	0%	26%	0%
	neu	6%	0%	3%	82%	0%	4%	6%		neu	6%	0%	3%	82%	0%	4%	6%
	joy	4%	6%	5%	10%	38%	34%	3%		joy	7%	4%	19%	30%	11%	30%	0%
	sur	4%	6%	4%	11%	12%	52%	12%		sur	11%	7%	4%	26%	4%	37%	11%
	dis	14%	20%	6%	12%	2%	22%	25%		dis	7%	26%	4%	30%	0%	22%	11%
Modalità di raccolta B		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta C		sad	ang	fea	neu	joy	sur	dis
	sad	56%	0%	7%	26%	0%	7%	4%		sad	75%	3%	5%	6%	4%	6%	3%
	ang	0%	56%	4%	15%	7%	15%	4%		ang	0%	64%	4%	5%	6%	18%	5%
	fea	30%	4%	33%	15%	11%	7%	0%		fea	26%	3%	38%	6%	2%	24%	2%
	neu	6%	0%	3%	82%	0%	4%	6%		neu	6%	0%	3%	82%	0%	4%	6%
	joy	4%	15%	0%	0%	44%	30%	7%		joy	3%	5%	3%	7%	44%	36%	3%
	sur	7%	4%	4%	11%	11%	44%	19%		sur	2%	6%	4%	7%	14%	57%	10%
	dis	15%	19%	7%	11%	4%	22%	22%		dis	16%	19%	6%	7%	3%	21%	29%

Tabella 19: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue francese

²⁹⁶ I valori di corretta identificazione sono riportati in diagonale per ciascuna delle modalità di raccolta con un ordine di lettura da sinistra a destra. Il totale per ciascuna riga può non corrispondere al 100% in quanto vengono riportati in tabella valori approssimati. Lo stesso vale per le tabelle che seguono.

INGLESE																	
Valori complessivi		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta A		sad	ang	fea	neu	joy	sur	dis
	sad	60%	1%	15%	13%	0%	3%	8%		sad	33%	4%	25%	21%	0%	0%	17%
	ang	3%	49%	8%	6%	4%	15%	16%		ang	13%	25%	8%	17%	4%	8%	25%
	fea	28%	1%	38%	13%	4%	4%	11%		fea	21%	4%	29%	21%	4%	4%	17%
	neu	31%	3%	2%	60%	1%	0%	2%		neu	31%	3%	2%	60%	1%	0%	2%
	joy	2%	6%	8%	14%	26%	37%	8%		joy	4%	8%	8%	33%	4%	42%	0%
	sur	8%	4%	9%	7%	14%	49%	9%		sur	8%	13%	8%	21%	8%	29%	13%
	dis	13%	10%	19%	8%	10%	17%	23%		dis	13%	21%	17%	13%	8%	25%	4%
Modalità di raccolta B		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta C		sad	ang	fea	neu	joy	sur	dis
	sad	67%	0%	13%	13%	0%	4%	4%		sad	66%	0%	14%	10%	0%	3%	7%
	ang	0%	63%	4%	8%	8%	13%	4%		ang	2%	51%	8%	2%	3%	17%	17%
	fea	4%	0%	50%	21%	8%	4%	13%		fea	35%	1%	38%	9%	3%	4%	9%
	neu	31%	3%	2%	60%	1%	0%	2%		neu	31%	3%	2%	60%	1%	0%	2%
	joy	0%	0%	13%	8%	17%	54%	8%		joy	2%	6%	7%	10%	33%	31%	9%
	sur	4%	8%	8%	4%	0%	67%	8%		sur	9%	1%	9%	4%	19%	49%	8%
dis	4%	8%	29%	4%	17%	17%	21%	dis	15%	7%	18%	8%	9%	15%	28%		

Tabella 20: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue inglese

TEDESCO																	
Valori complessivi		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta A		sad	ang	fea	neu	joy	sur	dis
	sad	66%	1%	10%	16%	1%	4%	4%		sad	58%	0%	3%	24%	0%	3%	12%
	ang	5%	59%	5%	10%	3%	10%	9%		ang	3%	42%	3%	24%	3%	3%	21%
	fea	17%	6%	40%	12%	2%	16%	7%		fea	9%	6%	27%	3%	3%	30%	21%
	neu	11%	2%	2%	83%	1%	2%	1%		neu	11%	2%	2%	83%	1%	2%	1%
	joy	3%	14%	6%	9%	35%	27%	7%		joy	3%	9%	15%	21%	24%	21%	6%
	sur	14%	10%	8%	11%	7%	40%	10%		sur	15%	12%	18%	15%	0%	24%	15%
	dis	15%	15%	13%	11%	8%	20%	18%		dis	18%	9%	18%	21%	0%	15%	18%
Modalità di raccolta B		sad	ang	fea	neu	joy	sur	dis	Modalità di raccolta C		sad	ang	fea	neu	joy	sur	dis
	sad	67%	0%	12%	9%	0%	6%	6%		sad	67%	1%	11%	15%	1%	4%	1%
	ang	15%	61%	3%	6%	0%	12%	3%		ang	3%	63%	5%	8%	3%	11%	7%
	fea	18%	6%	52%	12%	0%	9%	3%		fea	19%	5%	41%	14%	2%	14%	5%
	neu	11%	2%	2%	83%	1%	2%	1%		neu	11%	2%	2%	83%	1%	2%	1%
	joy	6%	15%	6%	9%	21%	30%	12%		joy	2%	15%	3%	6%	41%	27%	6%
	sur	21%	9%	6%	6%	3%	45%	9%		sur	11%	10%	6%	11%	10%	43%	9%
dis	21%	21%	9%	6%	12%	18%	12%	dis	6%	15%	15%	15%	9%	12%	27%		

Tabella 21: Matrici di confusione per il test di identificazione T1 per il sottocorpus mistilingue tedesco

Esaminando i valori complessivi di corretta identificazione, si rileva, citando Scherer, Johnstone, Klasmeyer (2003: 444), come le emozioni “*Sadness and anger are generally best recognized vocally, followed by fear*” con valori medi di corretta identificazione tra i più alti in tutte e quattro le lingue dopo le produzioni neutre, seguiti da gioia e disgusto (per le quali si rilevano, invece, le percentuali più basse). Ciò confermerebbe, quindi, quanto affermato da Johnstone, Scherer (2000), secondo i quali, il contesto dell’evoluzione nei termini proposti da Darwin avrebbe selezionato e favorito per alcune emozioni maggiori caratterizzazioni nella

sfera visiva, mentre per altre nella sfera acustica: per la rabbia e la paura sarebbe stata maggiormente sviluppata l'espressione vocale, perché gli antenati dell'uomo potessero avvertirsi e minacciarsi in modo esplicito anche a lunghe distanze; per emozioni quali disgusto e gioia sarebbero, invece, stati selezionati e favoriti sviluppi biologici relativi alla mimica facciale per far sì che, fra membri appartenenti alla stessa comunità, ci si potesse capire al volo (ciò spiegherebbe, in questo caso, le minori percentuali di corretto riconoscimento rilevate).

Al di là delle matrici di confusione, per quel che concerne il presente test di validazione sono state considerate come utili tutte quelle produzioni (*item*) in cui almeno 2 dei 3 giudici ascoltatori hanno correttamente identificato l'emozione intesa in fase di *encoding* secondo le procedure di raccolta di cui al § 9.3.

Lingua	Tipologia	Encoder	N° prod. utili	% prod. utili	% per tipologia	% per lingua
Italiano	naif	AR	18	45,0%	44,2%	55,2%
		DL	11	27,5%		
		GC	23	57,5%		
		GB	21	52,5%		
		NM	12	30,0%		
	SP	21	52,5%			
	attore	AS	28	70,0%	68,5%	
		EO	23	57,5%		
		MS	30	75,0%		
		MT	32	80,0%		
PM		24	60,0%			
Francese	naif	MC	19	47,5%	36,7%	48,1%
		JS	14	35,0%		
		AM	8	20,0%		
		BP	13	32,5%		
		PP	14	35,0%		
		TS	20	50,0%		
	attore	LL	29	72,5%	70,8%	
		HD	26	65,0%		
Inglese	naif	AF	14	35,0%	41,3%	43,8%
		IH	21	52,5%		
		IR	13	32,5%		
		JB	15	37,5%		
		MB	17	42,5%		
	MC	19	47,5%			
	attore	BM	13	32,5%	51,3%	
		CW	28	70,0%		
	Tedesco	naif	AN	13	32,5%	
AT			19	47,5%		
BG			12	30,0%		
SB			25	62,5%		
GS			16	40,0%		
ML			13	32,5%		
OK			15	37,5%		
PG			13	32,5%		
PM		31	77,5%			
attore		RP	24	60,0%	51,3%	
	EF	17	42,5%			

Tabella 22: Produzioni utili scaturite dal test di validazione T1 per lingua, tipologia soggetto ed encoder

Sulla base di tale filtro (2 su 3) sono state complessivamente ritenute utili rispettivamente il 55,2% delle produzioni per l'italiano, il 48,1% per il francese, il 43,8% per l'inglese e il 45% per il tedesco, la cui composizione in termini di produzioni per modalità di raccolta, risulta nelle seguenti proporzioni (vedi Figura 40):

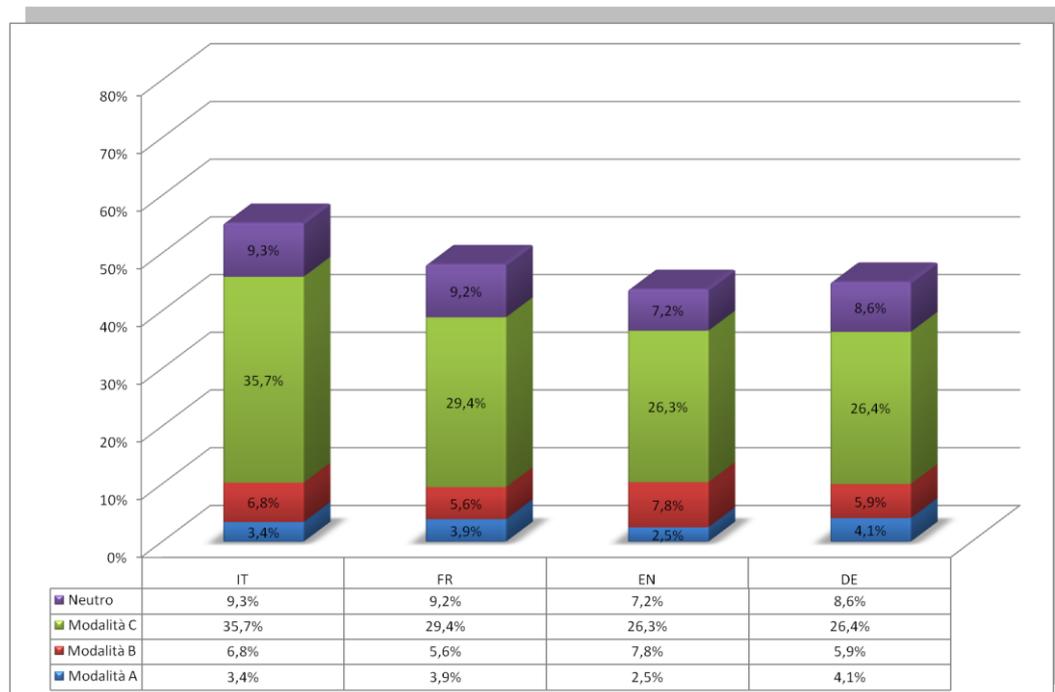


Figura 40: Consistenza del corpus per modalità di raccolta e per lingua in base al test di validazione T1

Se si esaminano, invece, i valori relativi a ciascuna modalità di raccolta (vedi Figura 41) scaturisce un altro dato importante.

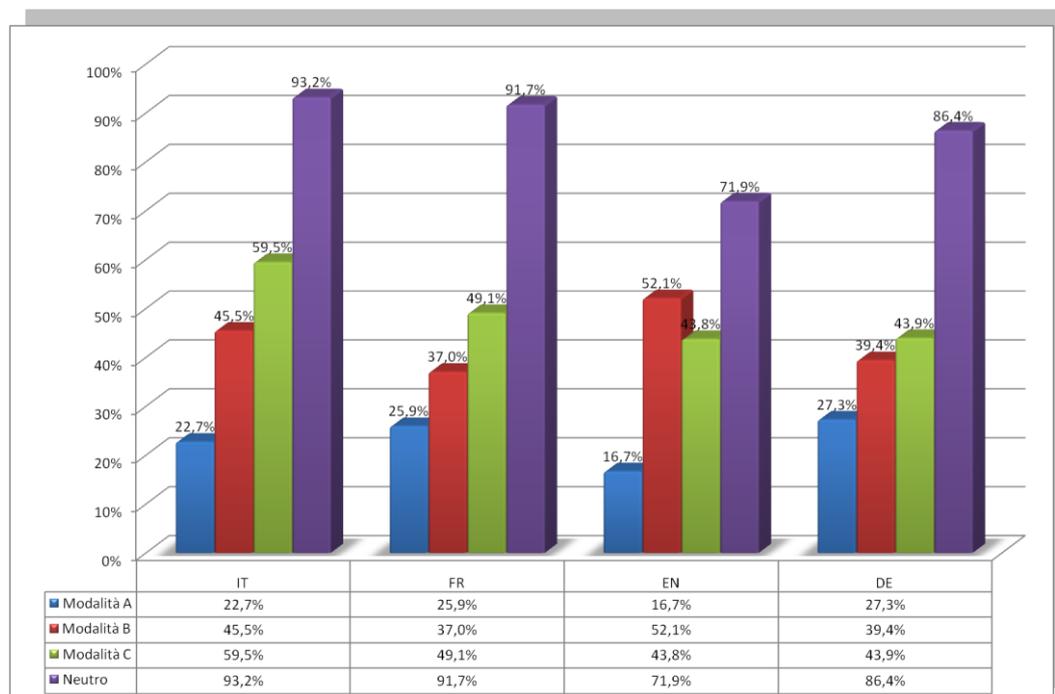


Figura 41: Materiale ritenuto utile per ciascuna lingua e per ciascuna modalità in base al test di validazione T1

Emerge, infatti, come i giudici ascoltatori di ciascuna lingua abbiano identificato con maggiore facilità le produzioni neutre seguite dalle produzioni raccolte in modalità isolata (modalità C), per finire con le più difficili da identificare nella modalità di raccolta A, confermando come la modalità A produca, date le sue caratteristiche, uno scarso coinvolgimento dell'*encoder* in quello che è rappresentato dallo scenario fornito.

Ad ogni modo, quali che siano le modalità di raccolta, si evince chiaramente come la validazione nei termini proposti da Magno Caldognetto (2002) sia non solo auspicabile, ma assolutamente necessaria al fine di stabilire con un buon grado di certezza ciò che si sta analizzando.

9.6.2 Test di validazione relativo alla rappresentatività delle emozioni (T2)

Per il test di validazione relativo alla rappresentatività delle emozioni presentate in riferimento all'emozione intesa in fase di *encoding*, prima di passare all'analisi dei risultati scaturiti da questa fase, si è proceduto alla verifica dell'attendibilità dei giudici ascoltatori in merito alla coerenza delle risposte fornite per uno stesso *item* (in questo caso rappresentati dai singoli *files*).

Poiché il test è stato costruito di modo che ciascun set di 40 produzioni di ciascun *encoder* venisse ascoltato da tre giudici ascoltatori (compreso l'*encoder* delle produzioni presentate), per ciascuna lingua e per ciascuno di tali set si proceduto al calcolo del coefficiente *alfa di Cronbach* (Cronbach, 1951) per verificare la consistenza dei dati sottoposti ad analisi. Il coefficiente *alfa di Cronbach* rappresenta una funzione tra il numero degli item del test N e l'intercorrelazione media tra gli item r secondo la seguente formula:

$$\alpha = \frac{N \times \bar{r}}{1 + (N - 1) \times \bar{r}}$$

Il calcolo di tale coefficiente è stato effettuato con la funzione RELIABILITY del software statistico SPSS.²⁹⁷

I valori medi rilevati per il coefficiente *alfa di Cronbach* sono: 0.672 per l'italiano, 0.640 per il francese, 0.558 per l'inglese e 0.636 per il tedesco. Va fatto notare, tuttavia, che il valore del coefficiente *alfa di Cronbach* cresce all'aumentare del numero degli *item* considerati: il numero ridotto di *item* (40) su cui il coefficiente è stato calcolato potrebbe in tal

²⁹⁷ Si veda a tal proposito la Tabella 17, dove il calcolo del coefficiente *alfa di Cronbach* è stato calcolato in modo trasversale a quello di generazione dei set di ascolto, ovvero da sinistra verso destra: ad es. con riferimento al set delle produzioni dell'*encoder* B in Tabella 17, il coefficiente *alfa di Cronbach* è stato calcolato sulle risposte fornite rispettivamente da A¹, B¹ e C¹ procedendo in modo analogo per tutte le altre.

caso aver influito sulla restituzione di valori relativamente bassi. Date le finalità qui perseguite e dal momento che lo scopo primario del T2 è quello di selezionare il materiale considerato più rappresentativo per le emozioni intese, i valori riscontrati, sebbene bassi, sono comunque da ritenersi sufficienti.²⁹⁸

Per il T2 è stata assunta come soglia di significatività un valore di rappresentatività pari o superiore a 2.3 per ciascun *item* dei test (su una scala di giudizio a 4 passi con un massimo di 4).

Lingua	Tipologia	Encoder	N° prod. utili	% prod. utili	% per tipologia	% per lingua		
Italiano	naif	AR	33	82,5%	66,3%	78,4%		
		DL	25	62,5%				
		GC	28	70,0%				
		GB	26	65,0%				
		NM	25	62,5%				
	SP	22	55,0%					
	attore	AS	38	95,0%	93,0%			
		EO	35	87,5%				
		MS	37	92,5%				
		MT	38	95,0%				
PM		38	95,0%					
Francese	naif	MC	26	65,0%	65,8%	75,3%		
		JS	28	70,0%				
		AM	23	57,5%				
		BP	25	62,5%				
		PP	20	50,0%				
	TS	36	90,0%					
	attore	LL	38	95,0%	94,2%			
		HD	38	95,0%				
		MO	37	92,5%				
	Inglese	naif	AF	21	52,5%		76,3%	81,3%
IH			26	65,0%				
IR			33	82,5%				
JB			34	85,0%				
MB			31	77,5%				
MC		38	95,0%					
attore		BM	39	97,5%	96,3%			
		CW	38	95,0%				
Tedesco		naif	AN	28	70,0%	74,4%	77,0%	
			AT	30	75,0%			
	BG		26	65,0%				
	SB		38	95,0%				
	GS		24	60,0%				
	ML		25	62,5%				
	OK		27	67,5%				
	PG		32	80,0%				
	PM	38	95,0%					
	attore	RP	36	90,0%	88,8%			
EF		35	87,5%					

Tabella 23: Produzioni utili scaturite dal test di validazione T2 per lingua, tipologia soggetto ed encoder

²⁹⁸ Come riportano Barbaranelli, D'Olimpio (2007: 241) non esistono regole statistiche per l'interpretazione del coefficiente di attendibilità, "ma si segue una regola pratica secondo la quale valori uguali almeno a .90 vengono considerati ottimi, valori compresi tra .80 e .90 molto buoni, valori compresi tra .70 e .80 buoni, valori compresi tra .60 e .70 sufficienti, valori inferiori a .60 inadeguati."

In base alla validazione effettuata con quello che è stato definito T2, è stato complessivamente ritenuto utile il 78,4% delle produzioni raccolte per il sotto corpus italiano, il 75,3% per quello francese, l'81,3% per quello inglese e il 77% per quello tedesco.

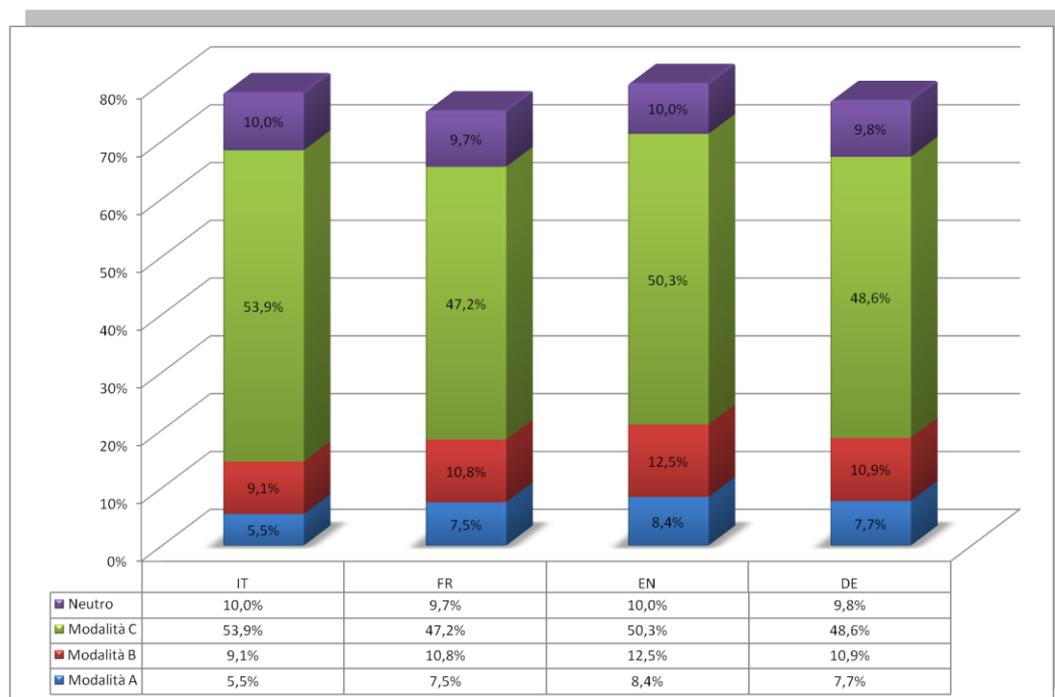


Figura 42: Consistenza del corpus per modalità di raccolta e lingua in base al test di validazione T2

La consistenza interna dei sotto *corpora* che compongono il corpus mistilingue qui raccolto in base alle modalità di raccolta più sopra esplicitate, risulta altrettanto simile nelle quattro lingue prese in considerazione. Si osservi, infine, il grafico di seguito riportato.

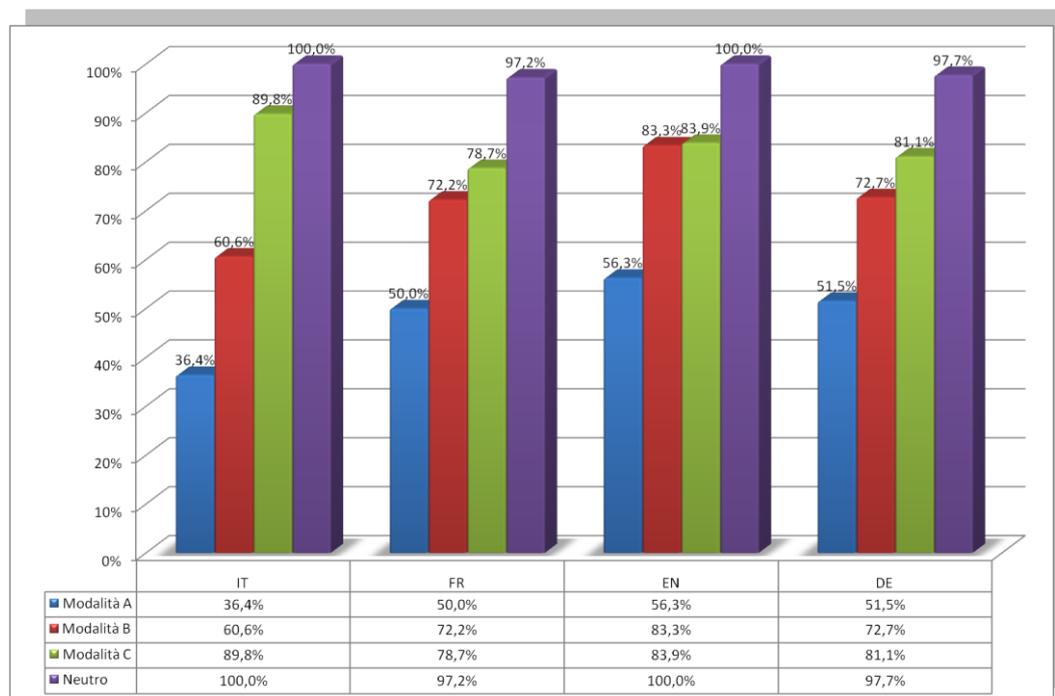


Figura 43: Materiale utile per ciascuna lingua e per ciascuna modalità di raccolta in base al test di validazione T2

Sebbene la parte più cospicua e presente all'interno dei vari sotto *corpora* (che compongono il corpus mistilingue) sia rappresentata in termini assoluti dalla modalità di raccolta C (modalità per la quale è stato possibile raccogliere più produzioni), è interessante rilevare come, in termini relativi, le produzioni appartenenti alla modalità Neutra e alla modalità di raccolta C siano le produzioni considerate dai giudici ascoltatori tra le più rappresentative (*cfr.* Figura 43).

9.6.3 Quale modalità di raccolta

Un altro obiettivo della presente ricerca consisteva nel verificare e valutare l'influenza del protocollo di elicitazione in rapporto alla riconoscibilità e alla rappresentatività delle produzioni emotive raccolte.

Dalle analisi sopra riportate emerge chiaramente come la modalità di raccolta C (che può essere considerata come raccolta di produzioni emotive "posate") sia di gran lunga la più efficace se si considera la quantità di materiale utile scaturita dalle procedure di validazione adottate in questa sede.

Sebbene le modalità di raccolta che sono state qui etichettate come modalità di raccolta A e B, restituiscano anch'esse produzioni riconoscibili e sufficientemente rappresentative delle emozioni intese, vi sono sostanzialmente due grossi limiti nel loro utilizzo: se, da un lato, il dispendio in termini di sforzi e di tempo nella raccolta del materiale sonoro diventa inevitabilmente estenuante, dall'altro si rischia di non disporre, a fine raccolta e a fine validazione, del materiale necessario allo svolgimento della ricerca avviata. Come si rileva dai dati riportati in Figura 41 per le due modalità A e B rispetto al test di validazione T1, e se si rapportano le percentuali al numero di produzioni emotive effettivamente utili (rispettivamente 15 su 66 e 30 su 66 per l'italiano, 14 su 54 e 20 su 54 per il francese, 8 su 48 e 25 su 48 per l'inglese e 18 su 66 e 26 su 66 per il tedesco), è facile intuire come non sia possibile rappresentare, per ciascun soggetto, un set completo composto di 6 produzioni emotive (ad es. per l'italiano per la modalità A sono state valutate utili solo 15 produzioni emotive su un totale di 66 relative agli 11 soggetti registrati).²⁹⁹

Va parimenti sottolineato come assai probabilmente la sequenza di acquisizione delle registrazioni, secondo le modalità a cui si è più volte accennato, caratterizzate da un livello di informazione e coinvolgimento del soggetto via via crescente, abbia aiutato i soggetti a

²⁹⁹ I dati qui riportati sono relativi alle produzioni utili ricavate sulla base di soglie di significatività individuate nei precedenti paragrafi.

focalizzarsi meglio sulle emozioni demandate nella modalità C. Da questo punto di vista la procedura adottata potrebbe addirittura essere considerata propedeutica per far esperire al soggetto le emozioni intese in ciascuno degli scenari presentati.

9.6.4 Soggetti *naif* o attori

Un altro aspetto qui indagato mirava a verificare e valutare l'influenza del protocollo di elicitazione in rapporto al soggetto registrato (*naif* vs. *attore*) valutandone la resa in termini di materiale utile. Senza ombra di dubbio l'utilizzo di *attori* produce sempre e comunque risultati migliori in termini di produzioni correttamente identificate e in termini di produzioni rappresentative.

Ne è un esempio il grafico riportato in Figura 44 relativo ai due test in cui si rileva come, a parità di condizioni (procedure e protocollo di raccolta), l'ago della bilancia penda complessivamente a favore degli *attori*, nonostante qualche trascurabile differenza tra le quattro lingue.

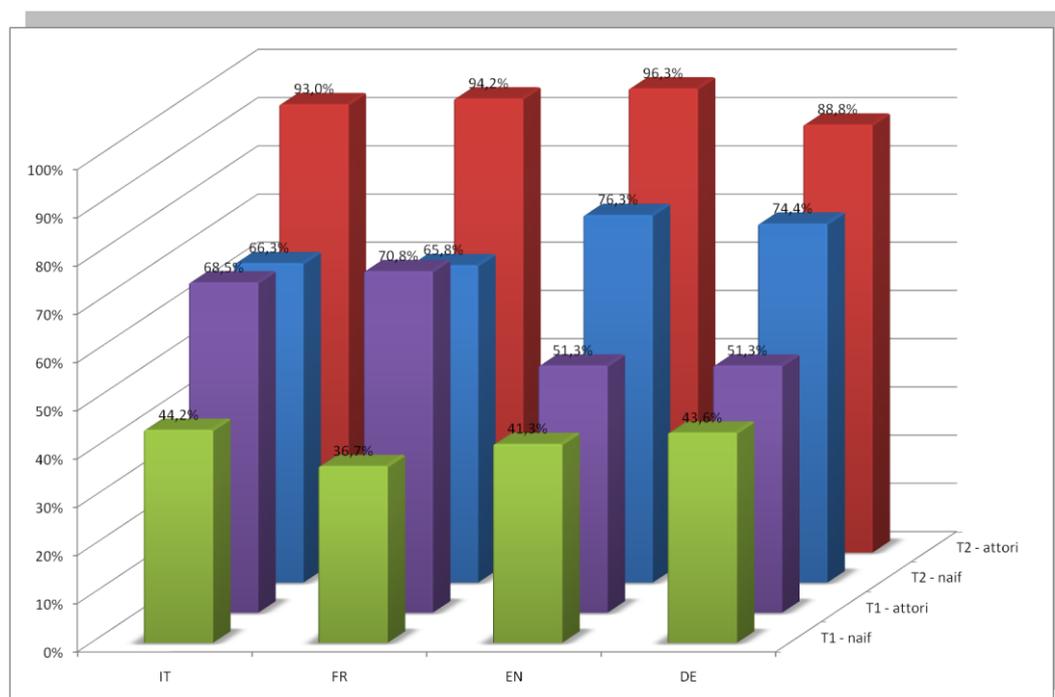


Figura 44: Produzioni ritenute utili per i due test di validazione con riferimento alle due categorie di encoder (*naif* vs. *attori*)

Nel caso di ricerche di tipo sperimentale sul parlato emotivo, l'utilizzo degli *attori* resta sicuramente una delle scelte più idonee, anche se i soggetti *naif* non sono affatto da disdegnare.

9.7 Conclusioni

In questo capitolo sono state motivate e descritte le scelte metodologiche e le procedure di raccolta del corpus mistilingue di parlato emotivo per l'italiano, il francese, l'inglese e il tedesco.

Sebbene i risultati qui riportati si riferiscano nello specifico alla validazione del corpus raccolto, operazione come più volte ribadito necessaria al prosieguo della ricerca intrapresa, è stato comunque possibile accertare e verificare una serie di presupposti già confermati in precedenti ricerche, come ad esempio l'abilità da parte di giudici ascoltatori a riconoscere emozioni vocali presentate nella propria lingua madre e la maggiore riconoscibilità a livello vocale di emozioni come rabbia e tristezza (Scherer, Johnstone, Klasmeyer, 2003).

A parità di condizioni, l'utilizzo di *attori* ha rivelato una maggiore resa in termini di materiale utile, nonostante non siano da disdegnare i soggetti *naif*.

Rispetto al protocollo di raccolta delle emozioni vocali qui adottato, va invece evidenziato come la produzione di frasi sulla base di scenari idonei (modalità di raccolta C nel presente lavoro) resti, al momento, una delle vie più percorribili nel caso di ricerche di tipo sperimentale (ancor più nel caso di ricerche di tipo cross-linguistico).

Mettendo a confronto i risultati della validazione effettuata, si rileva come la tipologia di test adottato, rispettivamente di identificazione delle emozioni nel T1 e di rappresentatività delle stesse nel caso di T2, possa diversamente discriminare tra produzioni utili e non per via della diversa difficoltà cognitiva del *task*.

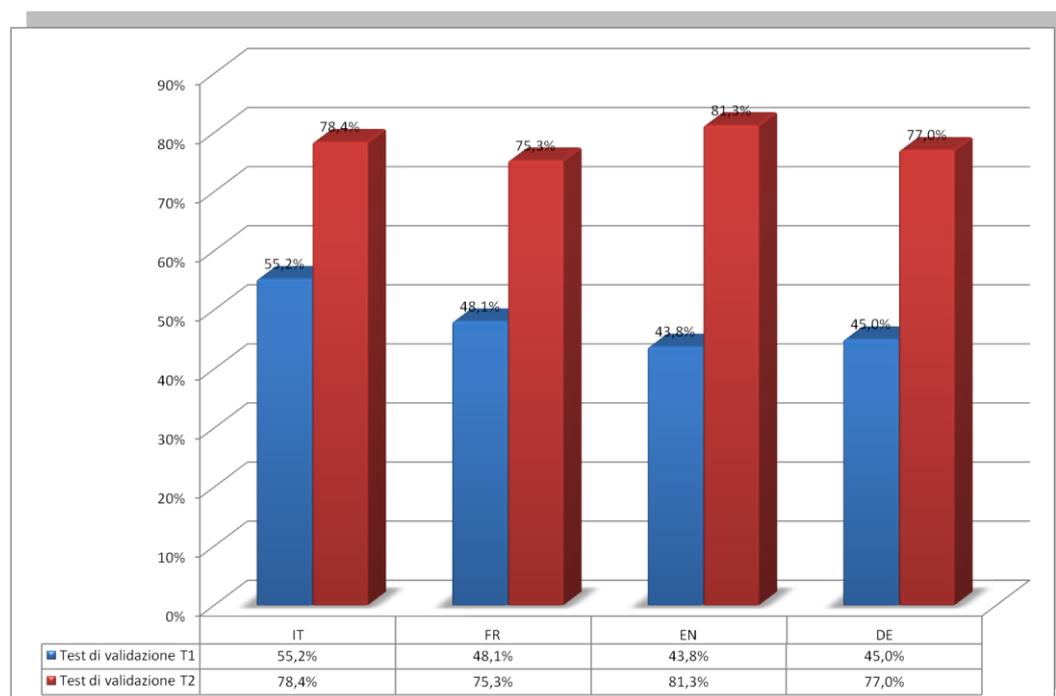


Figura 45: Confronto tra i due test di validazione (T1 e T2) con riferimento al materiale ritenuto utile per ciascuna lingua

Sebbene i due test risultino sufficientemente correlati, il diverso grado di difficoltà insito è senza dubbio da tenere in considerazione. Dal tipo di test e sulla base delle soglie di significatività individuate nel presente lavoro, dipende la caratterizzazione dei dati finali, come evidenziato, tra l'altro, nel grafico riportato in Figura 45.

In termini di capacità di discriminazione appare quindi evidente (vedi Figura 45) come sia sicuramente da prediligere (a scapito della quantità) un test di identificazione di emozioni che consenta di escludere tutte quelle produzioni che potrebbero creare confusione negli ascoltatori o nelle analisi acustiche che si intendono effettuare.

I dati presentati in questo capitolo, oltre a colmare un *vacuum* caratterizzato dall'assenza di un corpus emotivo mistilingue (contenente, tra le altre, anche la lingua italiana), costituiscono il punto di partenza dei prossimi due esperimenti all'oggetto della ricerca qui proposta, ovvero di:

- un test percettivo di tipo cross-linguistico in cui vengono presentate a soggetti di madrelingua italiana produzioni emotive nelle quattro lingue qui raccolte;
- un'analisi acustica delle produzioni somministrate nel predetto test percettivo volta alla caratterizzazione acustica delle emozioni nelle quattro lingue.

10 Step 2: Decoding cross-linguistico delle emozioni vocali

Per rispondere ai quesiti illustrati nel § 8.2, tra cui quelli di verificare l'abilità di soggetti italiani a decodificare emozioni espresse per mezzo della voce nella propria lingua madre (italiano) e in lingue diverse dalla propria (francese, inglese e tedesco) e di stabilire in che misura il grado di conoscenza e competenza della lingua possano influire sul riconoscimento delle produzioni emotive presentate, è stato predisposto un test di tipo percettivo da somministrare ad un campione di ascoltatori di madrelingua italiana.

10.1 La selezione degli stimoli emotivi

Gli stimoli utilizzati in questa fase sono stati opportunamente selezionati dal corpus di parlato emotivo mistilingue raccolto, descritto e validato nel precedente capitolo 9.

Partendo dai risultati della validazione effettuata con il test T1 (vedi Tabella 22) che, per il tipo di impostazione e per il tipo di task, ha rivelato un maggiore potere discriminante in termini di produzioni riconosciute come valide, per ciascuna lingua è stata effettuata una prima selezione che ha portato a ritenere gli encoder (un attore e un naif) per i quali è stato registrato il maggior numero di produzioni considerate utili (vedi tabella di seguito riportata).

Lingua	Tipologia	Soggetto	N° prod. utili	% prod. utili
Italiano	naif	GC	23	57,5%
	attore	MS	30	75,0%
Francese	naif	TS	20	50,0%
	attore	LL ³⁰⁰	29	72,5%
Inglese	naif	IH	21	52,5%
	attore	CW	28	70,0%
Tedesco	naif	PM	31	77,5%
	attore	RP	24	60,0%

Tabella 24: Selezione degli encoder con il maggior numero di produzioni ritenute utili

³⁰⁰ Stando a quanto riportato in Tabella 22, per la categoria degli attori francesi si sarebbe dovuto scegliere l'encoder MO (75%). Tuttavia, per una questione di preferenza e di gradimento soggettivo del timbro vocale si è optato per l'encoder LL.

Successivamente, per ciascuna delle voci selezionate (vedi Tabella 24) è stata scelta una produzione per ciascuna emozione (compresa la produzione neutra) adottando come criterio di selezione il tasso di riconoscimento più alto. A parità di punteggio riguardo il tasso di riconoscimento è stato adottato un ulteriore criterio di selezione optando per le produzioni relative alla modalità di raccolta C.³⁰¹

Encoder	Italiano (it)		Francese (fr)		Inglese (en)		Tedesco (de)	
	Naif (n)	Attore (a)	Naif (n)	Attore (a)	Naif (n)	Attore (a)	Naif (n)	Attore (a)
	GC	MS	TS	LL	IH	CW	PM	RP
Anger	c1	c4	c4	c4	c1	c1	c2	c4
Disgust	c1	c2	c1	b1	c1	c3	c1	c4
Fear	c2	c1	a1	c1	c2	c1	c4	c1
Joy	c2	b1	c4	c3	c3	c4	b1	c4
Neutral	3	4	3	1	4	1	3	3
Sadness	c1	c2	c1	c1	c3	c4	c4	c4
Surprise	c2	c2	c4	c4	b1	c2	c3	c1

Tabella 25: elenco delle produzioni selezionate per la somministrazione del test percettivo

Come si può dedurre dall'elenco delle produzioni in Tabella 25 in alcuni casi si sono avuti valori di riconoscimento più alti per produzioni raccolte nelle modalità A e B.

Le stesse produzioni sono state utilizzate oltre che per la somministrazione del test percettivo, di cui verranno fornite le caratteristiche nel prossimo §, anche per la successiva fase della ricerca riguardante l'analisi acustica delle produzioni emotive.

10.2 Preparazione del test percettivo

Poiché in fase di definizione della sperimentazione era stato deciso di somministrare il test percettivo via web per raggiungere il maggior numero possibile di soggetti, tale decisione ha notevolmente ridotto il campo delle possibilità di scelta tra i numerosissimi software attualmente disponibili.

In una fase esplorativa, alla ricerca di un software che consentisse di generare e di gestire un esperimento percettivo via web, si era inizialmente preferito utilizzare uno dei tanti software *open-source* resi disponibili nella rete dagli stessi autori che li avevano creati per specifici esperimenti scegliendo *SKATTA*³⁰², un software multiplatforma per la generazione e la somministrazione di test percettivi. Tuttavia, nonostante la sua semplicità d'uso, in corso

³⁰¹ In pochissimi casi, nel persistere delle condizioni di parità rispetto ai criteri di selezione adottati, si optato per la produzione avente il valore di rappresentatività del T2 più alto.

³⁰² Cfr. Bresin, R. & Lindberg, J. (2008), *SKATTA 1.1-beta5: Multiplatform system for the design and running of audio and visual perception tests*, [Computer program], retrieved from <http://sourceforge.net/projects/skatka/>.

d'opera si sono palesate alcune problematiche³⁰³ che, data l'impossibilità di un intervento risolutivo in tempi rapidi, ne hanno comportato l'abbandono a favore di un software di tipo commerciale.³⁰⁴

Per il suddetto motivo, e dopo averne appurato e verificato le potenzialità (ovvero la possibilità di generare test, quiz e questionari con la contestuale presentazione di materiali multimediali), la scelta si è orientata verso una piattaforma, *Questionmark™ Perception™*, già in uso e in licenza al Dipartimento di Linguistica dell'Università della Calabria. Si tratta di una piattaforma utilizzata prevalentemente per accertamenti di tipo didattico che consente di generare ed eseguire test, di sviluppare questionari a video ecc. che possono essere orientati al web per una piena automatizzazione e diffusione.



Figura 46: Caratteristiche di *Questionmark™ Perception™*³⁰⁵

Grazie alle potenzialità offerte dalla piattaforma e alla sua semplicità d'uso, il test percettivo è stato implementato in *Questionmark™ Perception™* per la somministrazione via web, sfruttando al massimo le caratteristiche del software e allo stesso tempo utilizzando un'interfaccia grafica facile ed intuitiva per la presentazione dell'intero test.

Il test implementato sulla piattaforma si compone di tre distinti blocchi consecutivi l'uno all'altro e ciascuno con caratteristiche proprie.

Nel primo blocco (*Blocco 1*) sono state fornite ai soggetti poche informazioni necessarie allo svolgimento del test con alcune raccomandazioni per un corretto svolgimento

³⁰³ Le problematiche riscontrate nel presente caso, e che interessano la maggior parte dei software *open-source* riguardano i seguenti aspetti:

- a) l'interfaccia utente per la generazione dell'esperimento è spesso poco intuitiva (se non addirittura assente) e segue spesso forme di *scripting* con specifici linguaggi di programmazione poco familiari;
- b) nascono spesso per specifiche esigenze di chi li ha creati, il che comporta l'assenza di manuali d'uso;
- c) poiché gli esperimenti somministrati con software di questo tipo vengono spesso eseguiti su una singola postazione, a volte manca del tutto una fase di *testing* per verificarne la compatibilità e il corretto funzionamento con i vari sistemi operativi e con i vari browser attualmente diffusi.

³⁰⁴ Al seguente link è disponibile una lunga lista dei software più utilizzati in questo ambito con caratteristiche e finalità varie: <http://www.psychology.org/links/Resources/Software/> (data consultazione: 24 febbraio 2009).

³⁰⁵ Per maggiori dettagli vedi <http://www.questionmark.com/us/perception/> (data consultazione: 2 marzo 2009).

dello stesso. Ai soggetti ascoltatori³⁰⁶ veniva inoltre chiesto di prestare attenzione a come le frasi venivano pronunciate piuttosto che al loro significato. Nella stessa pagina è stato inserito un modulo (*form*) per la raccolta di semplici informazioni di tipo anagrafico come nome, età, sesso e nazionalità (*cf.* schermata riportata in Appendice VII, Figura 72).

Il passaggio dal primo al secondo blocco (*Blocco 2*), attraverso un pulsante (*Submit*) presente nel primo blocco, dava inizio al test percettivo vero e proprio avente come oggetto emozioni vocali espresse a livello cross-linguistico (*cf.* schermata riportata in Appendice VII, Figura 73). In questa fase gli stimoli vengono presentati al soggetto singolarmente. Su ciascuna pagina e per ciascuno stimolo al soggetto vengono fornite:

1. l'istruzione riguardo il compito da svolgere;
2. un *player* con i comandi necessari alla riproduzione dello stimolo presentato;
3. la formulazione della domanda a cui rispondere con le sette possibili opzioni, che sono le stesse di quelle utilizzate nelle precedenti fasi di validazione degli stimoli, ovvero tristezza, collera, paura, neutro, gioia, sorpresa e disgusto.³⁰⁷

In questa fase sono stati complessivamente presentati al soggetto 56 stimoli vocali emotivi (per le modalità di selezione di tali stimoli si veda quanto riportato nel § 10.1). Il passaggio da uno stimolo all'altro avveniva su richiesta del soggetto attraverso un apposito pulsante (*Next*) in fondo alla stessa pagina, con la sola possibilità di passare allo stimolo successivo.³⁰⁸ Al soggetto era data comunque facoltà di riascoltare più volte lo stimolo presentato attraverso l'utilizzo dei comandi disponibili nel *player*. Gli stimoli sono stati presentati ai soggetti in modalità *random*. Solo dopo aver visualizzato e risposto al 56^{esimo} stimolo al soggetto si rendeva visibile un pulsante (*Proceed*) attraverso il quale inviare le risposte e passare alla pagina successiva relativa al terzo blocco (*Blocco 3*).

Nel *Blocco 3* (*cf.* schermata riportata in Appendice VII, Figura 74) sono state inserite una serie di domande, consecutive l'una all'altra, volte a verificare alcuni presupposti a cui si è accennato nel § 8.2.

Per verificare se alcune impressioni personali da parte dei soggetti ascoltatori in merito allo svolgimento del test trovino effettivamente riscontro nei valori di riconoscimento relativi agli stimoli somministrati, a ciascun soggetto sono stati formulati due quesiti:

³⁰⁶ Nel prosieguo si farà indifferentemente uso dei termini “soggetto ascoltatore” e “rispondente” come sinonimi.

³⁰⁷ Trattandosi di una domanda a risposta multipla era consentita una sola selezione e quindi una sola risposta.

³⁰⁸ Non era pertanto consentito tornare indietro per riascoltare precedenti stimoli o modificare le risposte date per ciascuno di essi.

1. “Per quale lingua hai avuto più difficoltà a riconoscere le emozioni?”, alla quale il soggetto poteva rispondere scegliendo una ed una sola delle quattro opzioni possibili (italiano, francese, inglese o tedesco);
2. “A prescindere dalla lingua, indica quali sono, secondo te, le due emozioni più difficili da riconoscere.”, in questo caso al soggetto ascoltatore era richiesto di indicare due delle sette emozioni in elenco.³⁰⁹

Per valutare, invece, in che misura il grado di conoscenza e competenza della lingua presentata possa influire sul riconoscimento delle produzioni emotive presentate, ovvero se a competenze maggiori corrisponda un numero maggiore di stimoli correttamente riconosciuti, ai soggetti ascoltatori sono state formulate le seguenti domande:

1. “Indica il tuo grado di competenza nella lingua italiana.”, a cui il soggetto ascoltatore poteva rispondere con una delle seguenti opzioni: *nessuna, apprendista, scolastica, universitaria, avanzata, madrelingua*; la stessa domanda è stata presentata nei medesimi termini anche per le lingue straniere (francese, inglese e tedesco);³¹⁰
2. “Sei in possesso di qualche certificazione (ad. es.: TOEFL, Trinity, DALF/DELF, ecc) per qualcuna delle lingue che qui hai ascoltato?”, in tal caso, qualora la risposta fosse stata affermativa al soggetto veniva anche chiesto di indicare quali.

Infine, si è ritenuto interessante ed utile anche una domanda del tipo “Hai qualche commento o osservazione da fare sull'esperimento a cui hai appena partecipato?” per dare al soggetto ascoltatore la possibilità di esprimere qualsiasi tipo di commento o impressione in un apposito campo note.

Anche in questo caso, come per le precedenti schermate, dopo aver risposto alle domande contenute nel *form* il soggetto era tenuto ad inviare le risposte fornite premendo il pulsante *Submit* situato in fondo alla pagina. Una schermata finale di chiusura, con cui si ringraziava il soggetto per la sua partecipazione alla sperimentazione comunicava, altresì, che il test era concluso.

³⁰⁹ Sebbene fosse richiesto di indicare soltanto due emozioni, qualcuno dei soggetti ne ha indicato solo una mentre altri ancora ne hanno indicato tre e in pochissimi casi quattro.

³¹⁰ Per quel che concerne le risposte possibili si è tentato di prevedere i vari livelli di competenza della lingua sulla base dei livelli di apprendimento delle lingue straniere. La domanda relativa all'italiano è stata inserita come elemento di controllo per escludere eventuali soggetti che non fossero di madrelingua italiana (non è infatti detto che un soggetto avente nazionalità italiana sia necessariamente di madrelingua italiana) e per dare ai soggetti un elemento di riferimento sui livelli di competenza riportati.

10.3 Somministrazione del test

Come già detto il test è stato interamente somministrato via web avvalendosi alla piattaforma *Questionmark™ Perception™* in uso e in licenza al Dipartimento di Linguistica dell'Università della Calabria.

Per il reclutamento dei soggetti si è fatto ricorso a comunicazioni inoltrate via mail. Oltre alla richiesta di collaborazione e partecipazione ai propri contatti, è stata data pubblicità dell'esperimento anche sulla piattaforma di *e-learning Moodle* in uso presso il Dipartimento di Linguistica dell'Università della Calabria³¹¹ e all'interno della *mailing-list* dell' AISV (Associazione Italiana di Scienze della Voce) pregando i suoi soci ad inoltrare, a loro volta, la richiesta di collaborazione ai loro contatti.

Nella richiesta inviata via mail, con cui si invitavano i soggetti a collaborare all'esperimento, venivano contestualmente fornite sommarie indicazioni sulla ricerca condotta e sugli obiettivi perseguiti, informazioni e requisiti tecnici necessari allo svolgimento dell'esperimento e il *link* al quale accedere per prendere parte alla sperimentazione.³¹²

In questo modo è stato possibile raggiungere e reperire, su base volontaria, un numero elevato di partecipanti su tutto il territorio nazionale.

Una volta dato avvio alla sperimentazione con l'attivazione della procedura di somministrazione dell'esperimento percettivo descritto nel precedente paragrafo, la stessa è rimasta attiva per un periodo di ca. 3 mesi, durante il quale i soggetti reclutati si sono collegati alla piattaforma per l'esecuzione del test.

L'intero test (comprensivo dei tre blocchi più sopra illustrati) è stato somministrato senza alcuna restrizione di tempo. È stata, quindi, lasciata sufficiente libertà ai soggetti ascoltatori coinvolti, offrendo loro la possibilità di ascoltare gli stimoli presentati all'interno del *Blocco 2* tante volte quanto lo ritenessero necessario.

Per tutta la fase di somministrazione del test si è reso necessario un costante monitoraggio dell'intera procedura al fine di prevenire eventuali malfunzionamenti che avrebbero impedito un corretto svolgimento o completamento delle prove percettive da parte dei soggetti ascoltatori.

³¹¹ La piattaforma di *e-learning* è attualmente utilizzata da diversi docenti e studenti.

³¹² Per maggiori dettagli si veda la mail riportata in Appendice IX. Tra le informazioni e i requisiti di tipo tecnico ai soggetti veniva chiesto di svolgere il test con l'ausilio di una cuffia.

10.4 Il campione dei soggetti ascoltatori

Poiché la comunicazione inoltrata, nello specifico, alla *mailing-list* AISV ha raggiunto anche soggetti stranieri e sebbene questi fossero competenti nella lingua italiana, per uniformare il campione dei soggetti ascoltatori si è deciso di escludere i loro risultati dalla presente indagine.³¹³ Data questa preliminare operazione di “filtraggio” il campione dei soggetti ascoltatori a cui è stato somministrato l’esperimento risulta costituito da 100 soggetti (56 di sesso femminile e 44 di sesso maschile) tutti di nazionalità e di madre lingua italiana.

L’età dei soggetti ascoltatori è compresa tra i 20 e i 62 anni per i soggetti di sesso femminile e tra i 20 e i 53 per quelli di sesso maschile. In Tabella 26 viene riportato il riepilogo delle statistiche relative all’età dei soggetti ascoltatori divisi per sesso.

Sesso	N	Età					
		Minimo	Massimo	Intervallo	Media	Dev. std.	Mediana
Femmina	56	20	62	42	30,00	7,804	28,50
Maschio	44	20	53	33	33,82	8,686	32,50
Totale	100	20	62	42	31,68	8,381	30,50

Tabella 26: Riepilogo statistiche età campione soggetti ascoltatori

Nella Figura 47 di seguito riportata si fornisce una rappresentazione grafica dei dati riportati in Tabella 26 che consente di meglio apprezzarne la distribuzione con riferimento all’età dei soggetti ascoltatori divisi, anche in questo caso, per sesso.

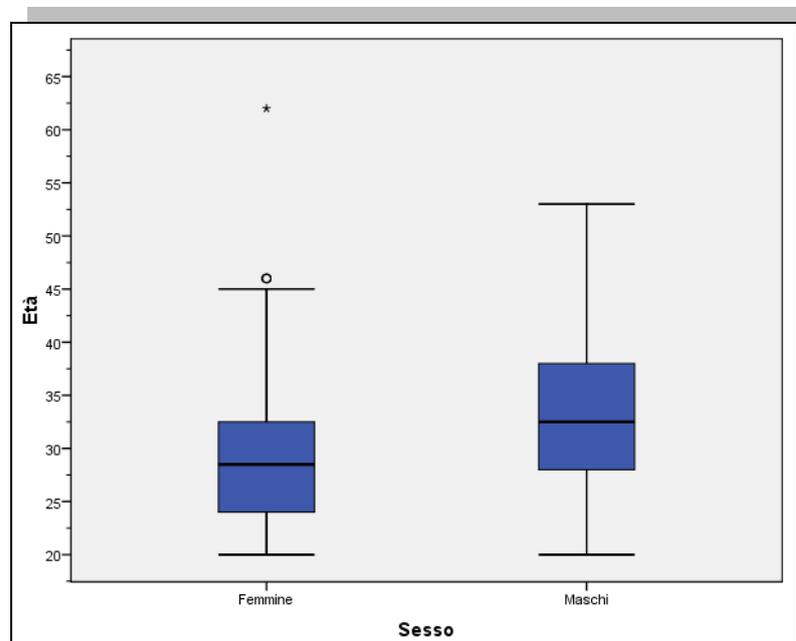


Figura 47: Distribuzione del campione dei soggetti ascoltatori per età e per sesso

³¹³ L’esclusione è stata possibile grazie all’indicazione della nazionalità all’interno del *Blocco 1* e dell’informazione relativa alla competenza nella lingua italiana contenuta nel *Blocco 3*. Sono stati quindi esclusi tutti quei soggetti (complessivamente 2 di sesso maschile e 4 di sesso femminile) la cui nazionalità non risultava essere italiana.

10.5 Risultati dell'esperimento e discussione dati

Si passerà ora all'analisi dei risultati scaturiti dall'attuazione del secondo *step* della ricerca che qui è stato descritto, l'esportazione dei quali si è resa possibile soltanto alla chiusura della procedura di somministrazione del test percettivo avvenuta, come detto, dopo circa 3 mesi dalla sua attivazione, ovvero al superamento della soglia dei 100 rispondenti.

10.5.1 Tempo utilizzato dai rispondenti per lo svolgimento del test

L'intero test (comprensivo dei tre blocchi più sopra illustrati) è stato somministrato senza alcuna restrizione di tempo. È stata, quindi, lasciata sufficiente libertà ai soggetti ascoltatori coinvolti, offrendo loro la possibilità di ascoltare gli stimoli presentati all'interno del *Blocco 2* tante volte quanto lo ritenevano necessario.

Dalle informazioni riportate in Tabella 27 si evince come, nel complesso, vi sia stata una certa uniformità nel tempo speso da parte dei rispondenti per il completamento dell'intero test percettivo loro somministrato, tempo che si aggira in media attorno ai 17 minuti.

Tempo impiegato (hh.mm.ss)		
N	Validi	100
	Mancanti	0
Media		00.17.14
Mediana		00.14.40
Deviazione std.		00.08.15
Intervallo		00.51.23
Minimo		00.07.06
Massimo		00.58.29
Percentili	10	00.10.28
	20	00.11.39
	25	00.12.13
	30	00.12.36
	40	00.13.30
	50	00.14.40
	60	00.16.07
	70	00.18.31
	75	00.19.25
	80	00.23.05
	90	00.26.30

Tabella 27: Riepilogo statistico del tempo impiegato dagli ascoltatori per il completamento dell'intero test

Osservando la Figura 48, in cui viene fornita una rappresentazione grafica delle informazioni riportate in Tabella 27 emerge, inoltre, come solo in pochissimi casi (3 in tutto) il tempo medio impiegato dai soggetti ascoltatori ha superato la mezz'ora di tempo.³¹⁴

³¹⁴ Le spiegazioni potrebbero essere in tal caso due: 1) il soggetto ha ascoltato più volte gli stimoli presentati, il che gli ha richiesto il orientativamente il doppio del tempo impiegato da tutti gli altri, oppure 2) il soggetto si è preso una o più pause durante lo svolgimento del test, il che ha fatto registrare tempi di svolgimento maggiori.

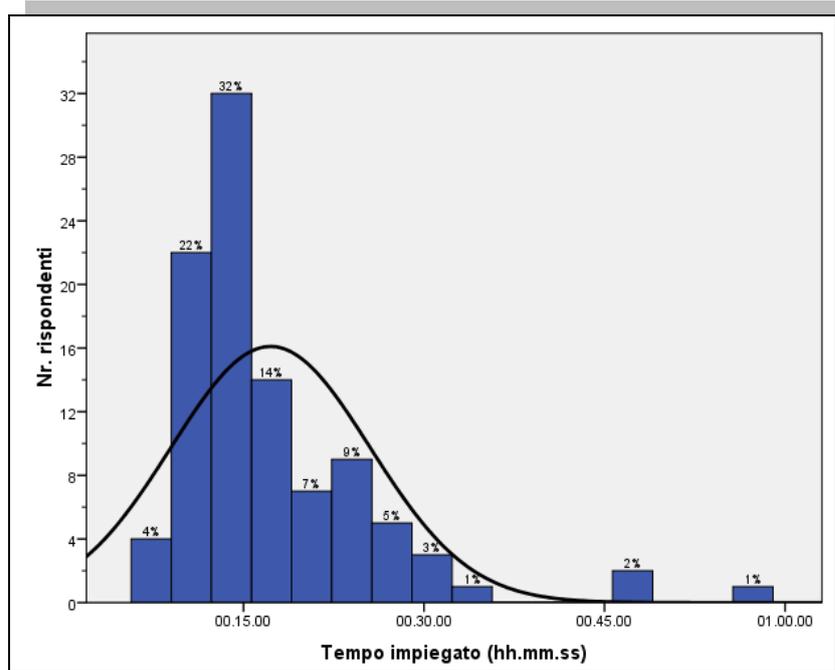


Figura 48: Distribuzione di frequenza in funzione del tempo impiegato dagli ascoltatori per il completamento del test

10.5.2 Decoding di emozioni vocali a livello cross-linguistico

Da una prima sommaria analisi dei risultati del test somministrato a 100 soggetti ascoltatori di madrelingua italiana, emerge come vi sia, da parte di questi ultimi, una maggiore abilità nel riconoscere produzioni vocali emotive nella propria lingua madre (*in-group advantage*), piuttosto che in lingue diverse dalla propria.

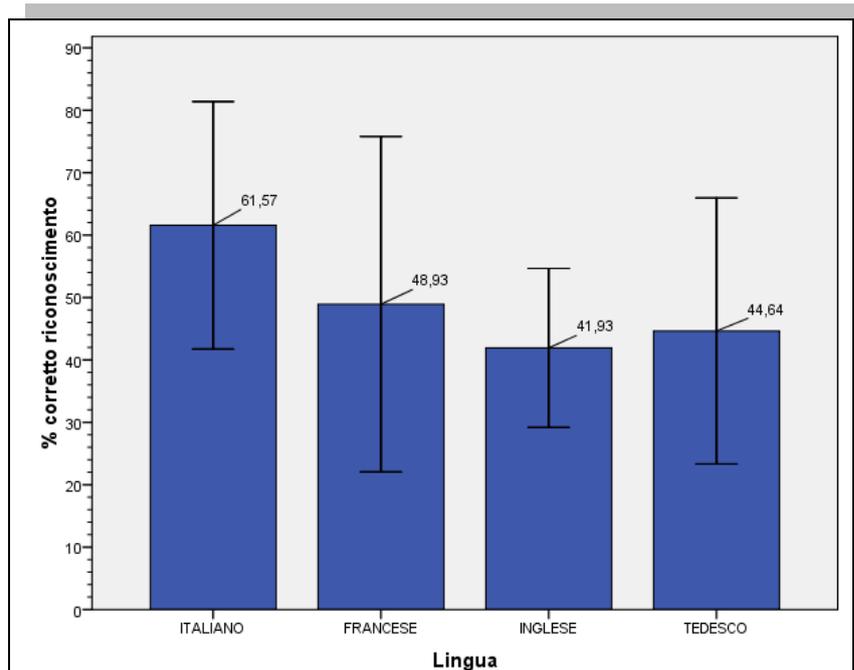


Figura 49: Percentuale di stimoli correttamente riconosciuti da parte di soggetti ascoltatori di madrelingua italiana per ciascuna delle lingue loro presentate

Esaminando il grafico riportato in Figura 49, si noter , infatti, come le medie percentuali di corretto riconoscimento oscillano da un 61,57% per stimoli presentati in lingua italiana, ad un 48,93%, 44,64% e 41,93% per stimoli presentati, rispettivamente, in lingua francese, tedesca e inglese. Il dato   sicuramente in linea con i risultati di precedenti ricerche condotte in questo ambito (*cf.* § 7.2) nelle quali si era gi  constatata una certa abilit  da parte dei soggetti a meglio riconoscere produzioni emotive nella propria lingua madre, abilit  definita appunto in termini di *in-group advantage*. Questo assunto, inoltre, trova ulteriore conferma nei dati in Tabella 28, di seguito riportata, nella quale vengono riassunte in una matrice di confusione le risposte dei 100 rispondenti a ciascuno degli stimoli presentati e qui raggruppati per lingua.

Stimolo presentato		Risposta							
Lingua	Emozione	<i>sad</i>	<i>ang</i>	<i>fea</i>	<i>neu</i>	<i>joy</i>	<i>sur</i>	<i>dis</i>	<i>n/a</i> ³¹⁵
Italiano	<i>sad</i>	83	0	2	9	0	4	3	0
	<i>ang</i>	2	82	1	3	0	4	8	1
	<i>fea</i>	26	2	55	2	1	11	4	0
	<i>neu</i>	7	4	1	78	0	5	7	1
	<i>joy</i>	6	4	3	1	51	30	6	0
	<i>sur</i>	13	5	2	15	0	54	12	1
	<i>dis</i>	23	16	3	9	2	19	30	0
Francese	<i>sad</i>	82	1	7	3	1	4	3	2
	<i>ang</i>	1	20	0	8	21	37	13	2
	<i>fea</i>	8	1	44	25	1	18	3	2
	<i>neu</i>	14	0	1	80	1	3	1	2
	<i>joy</i>	1	1	2	4	51	43	0	0
	<i>sur</i>	5	0	1	26	10	56	3	1
	<i>dis</i>	14	1	0	26	7	40	12	2
Inglese	<i>sad</i>	50	2	11	18	0	8	12	1
	<i>ang</i>	0	59	3	6	1	14	17	2
	<i>fea</i>	22	1	38	20	0	13	6	2
	<i>neu</i>	36	1	5	50	0	4	5	1
	<i>joy</i>	1	7	2	2	19	58	13	0
	<i>sur</i>	5	12	9	14	3	40	17	2
	<i>dis</i>	1	44	1	2	1	14	39	0
Tedesco	<i>sad</i>	66	2	13	14	0	1	4	1
	<i>ang</i>	4	40	4	15	3	20	16	0
	<i>fea</i>	21	17	36	6	4	12	6	1
	<i>neu</i>	9	2	1	83	0	2	3	2
	<i>joy</i>	1	26	5	7	30	26	5	1
	<i>sur</i>	3	18	3	21	8	30	18	1
	<i>dis</i>	10	18	10	23	3	8	29	1

Tabella 28: Matrice di confusione riportante le risposte agli stimoli presentati suddivisi per lingua

³¹⁵ Nel rispondere agli stimoli presentati   accaduto che qualcuno dei soggetti cliccasse erroneamente sul pulsante *Next* senza prima aver fornito una risposta allo stimolo ascoltato. Con “n/a” (*not available*) vengono pertanto conteggiati, nel prosieguo dell’esposizione, i casi di mancata risposta agli stimoli presentati, agevolando al contempo il computo delle percentuali.

Oltre a confermare la presenza di un *in-group advantage*, le risposte riportate in Tabella 28 sono coerenti con Scherer *et al.* (2003: 444), a cui si è più volte accennato, che afferma come le emozioni di tristezza e rabbia sono in genere le emozioni più riconosciute, seguite da paura. Ciò è vero se si considerano le emozioni espresse, in questo caso, nella stessa lingua dei rispondenti (italiano) e in lingua tedesca, con percentuali rispettivamente dell'83%, dell'82% e del 55% per l'italiano, e del 66%, del 40% e del 36% per il tedesco. Per le altre due lingue, inglese e francese, si riscontrano, invece, una serie di incongruenze.

Prima di procedere nell'analisi, sebbene la Tabella 28 offra una visione complessiva dei risultati del test percettivo, occorre aprire una breve parentesi per mettere in evidenza un elemento, tutt'altro che trascurabile, che caratterizza la ricerca e i dati qui presentati.

Se si analizza visivamente il grafico in Figura 49 e lo si accosta ai dati riportati in Tabella 29, si rileva come vi sia per tutte e quattro le lingue una deviazione standard di corretto riconoscimento piuttosto elevata (mediamente del 20% ca.).

Descrittive				
	Lingua		Statistica	Errore std.
	% corretto riconoscimento	Italiano	Media	61,5714
Mediana			54,5000	
Deviazione std.			19,81882	
Minimo			29,50	
Massimo			82,50	
Francese		Media	48,9286	10,15501
		Mediana	50,5000	
		Deviazione std.	26,86762	
		Minimo	12,00	
		Massimo	81,50	
Inglese		Media	41,9286	4,81494
		Mediana	39,5000	
		Deviazione std.	12,73914	
		Minimo	19,00	
		Massimo	59,00	
Tedesco		Media	44,6429	8,05614
		Mediana	35,5000	
		Deviazione std.	21,31454	
		Minimo	29,00	
		Massimo	83,00	

Tabella 29: Riepilogo statistiche corretto riconoscimento delle emozioni vocali presentate ai soggetti ascoltatori per ciascuna lingua

A falsare i dati di corretto riconoscimento, se così si può affermare, sono le produzioni emotive raccolte con i soggetti *naïf*, le cui produzioni sono state riconosciute con percentuali decisamente più basse rispetto a quelle prodotte da soggetti *attori* (vedi Tabella 30).

Durante le fasi di validazione delle produzioni emotive raccolte (*cfr.* § 9.6.4) prima, e nella formulazione delle ipotesi per questa fase della ricerca dopo, era infatti stata ipotizzata la

presenza di differenze di riconoscimento riconducibili a produzioni realizzate da *attori* o da soggetti *naif*.

valori medi corretto riconoscimento (%)			
	<i>attori</i>	<i>naif</i>	<i>complessivo</i>
<i>Italiano</i>	65,86	57,29	61,57
<i>Francese</i>	59,43	38,43	48,93
<i>Inglese</i>	47,57	36,29	41,93
<i>Tedesco</i>	51,43	37,86	44,64

Tabella 30: Valori medi di corretto riconoscimento per ciascuna lingua per stimoli prodotti rispettivamente da soggetti *attori* e soggetti *naif*

Infatti, rappresentando graficamente i dati in Tabella 30, emerge chiaramente come le percentuali di corretto riconoscimento totali risentano dei valori di corretto riconoscimento riferiti a stimoli prodotti in tutte e quattro le lingue da soggetti *naif*. Si assiste, di fatto, ad una oscillazione del 7,5% ca. tra stimoli prodotti da soggetti *attori* e stimoli prodotti da soggetti *naif* per la lingua italiana, oscillazioni ancor più evidenti ed apprezzabili nel caso di lingue diverse da quella dei soggetti ascoltatori (vedi Figura 50).

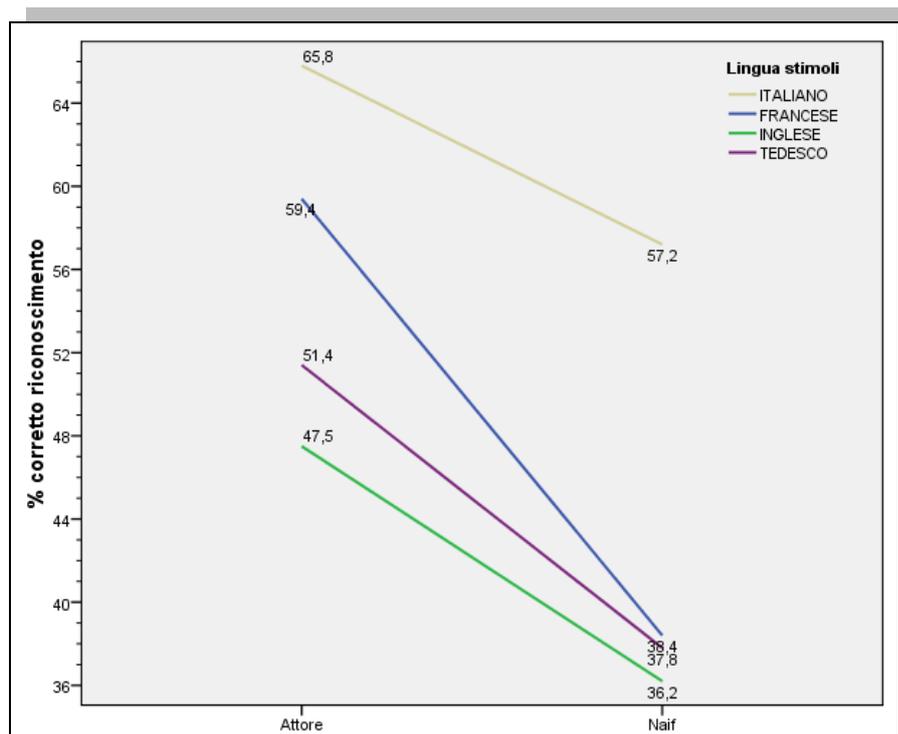


Figura 50: Differenze nel riconoscimento corretto di stimoli prodotti da *attori* e soggetti *naif*

Alla luce di quanto detto, prendendo a tal proposito in considerazione i valori di precedenti ricerche che hanno utilizzato soggetti *attori*, si può facilmente constatare come il valore del 65,8% relativo all'*in-group advantage* nella presente indagine sia coerente, ad esempio, con quanto riportato da Pell *et al.* (2009) per stimoli presentati nella stessa lingua

del *decoder* (64%), o ancora con i valori riportati da Graham *et al.* (2001), Scherer *et al.* (2001) e Van Bezooijen *et al.* (1983), rispettivamente del 59%, del 74% e del 66%.

Allo stesso tempo, nonostante le differenze di riconoscimento da parte dei soggetti ascoltatori dovute alla diversa abilità dei soggetti *attori* e dei soggetti *naif* nel produrre stimoli vocali emotivi, i valori medi di corretto riconoscimento si collocano sensibilmente al di sopra della soglia di riconoscimento data, nella presente indagine, dalla pura casualità.³¹⁶

Si può affermare, inoltre, come i dati appena presentati confermano, da questo punto di vista, quanto emerso, ad esempio, dallo studio di Wilting *et. al* (2006) in cui era stata evidenziata una maggiore riconoscibilità delle emozioni prodotte da soggetti *attori*.

È questa una delle ragioni che ha spinto e spinge la maggior parte dei ricercatori all'utilizzo di attori per la raccolta di emozioni vocali da utilizzare per la somministrazione di test di tipo percettivo.

Chiudendo questa breve parentesi si riscontra, in poche parole, una sostanziale differenza nel riconoscimento corretto degli stimoli a seconda che si tratti di stimoli prodotti da *encoder* in veste di soggetti *attori* o di *encoder* in veste di soggetti *naif* (*cfr.* Figura 50 per un riscontro visivo immediato). Per questo motivo, nel prosieguo delle analisi, si ritiene opportuno trattare i dati in modo disgiunto, separando le risposte dei soggetti ascoltatori in base agli *encoder* che hanno prodotto gli stimoli. Si farà perciò riferimento alla Tabella 31 di seguito presentata, sulla quale sono stati evidenziati in “**grassetto corsivo**” i casi più significativi.

Da una prima analisi visiva delle risposte in Tabella 31, se si confrontano, per ciascuna lingua, le risposte a stimoli prodotti da *attori* o soggetti *naif*, si evidenzia una certa sistematicità nella confusione di alcune emozioni: per citarne qualcuno, si veda ad esempio per il francese, come l'emozione di rabbia venga confusa in modo sistematico con le emozioni di gioia e sorpresa sia nel caso di stimoli prodotti da un *attore* sia nel caso di stimoli prodotti da un *non-attore*.

³¹⁶ Soglia che si rammenta essere del 14,3% ca. nel test qui somministrato che offriva, per la risposta, sette opzioni possibili.

Stimolo presentato		Risposta															
		Encoder = Attore								Encoder = Naif							
Lingua	Emozione	sad	ang	fea	neu	joy	sur	dis	n/a	sad	ang	fea	neu	joy	sur	dis	n/a
Italiano	sad	91	0	4	1	0	4	0	0	74	0	0	16	0	4	6	0
	ang	0	97	1	0	0	0	1	1	3	67	0	6	0	8	15	1
	fea	14	2	68	1	0	11	4	0	38	1	41	3	2	11	4	0
	neu	7	3	0	73	0	5	12	0	7	4	1	82	0	4	1	1
	joy	10	7	5	2	29	36	11	0	2	1	0	0	73	24	0	0
	sur	14	3	2	12	0	56	13	0	11	6	2	18	0	52	10	1
	dis	9	27	2	1	0	14	47	0	37	4	3	17	4	23	12	0
Francese	sad	80	1	9	4	1	3	2	0	83	0	4	2	0	4	4	3
	ang	0	22	0	1	34	37	4	2	2	17	0	15	7	36	21	2
	fea	1	1	86	3	0	5	2	2	14	1	1	47	1	31	4	1
	neu	9	0	0	86	1	3	1	0	18	0	2	74	0	2	1	3
	joy	2	0	3	0	72	23	0	0	0	1	0	7	29	63	0	0
	sur	5	0	1	28	16	47	3	0	4	0	1	23	4	64	3	1
	dis	21	1	0	30	5	18	23	2	6	1	0	21	8	62	1	1
Inglese	sad	55	1	17	10	0	7	10	0	44	3	4	25	0	9	13	2
	ang	0	78	6	0	1	3	11	1	0	40	0	11	0	25	22	2
	fea	10	0	71	0	0	15	2	2	33	1	4	40	0	11	9	2
	neu	41	1	7	38	0	8	5	0	30	0	3	62	0	0	4	1
	joy	2	11	4	0	8	61	14	0	0	2	0	3	30	54	11	0
	sur	5	18	16	2	2	39	17	1	4	5	2	26	4	40	17	2
	dis	1	46	0	1	1	7	44	0	0	41	1	2	1	21	34	0
Tedesco	sad	67	2	18	7	0	2	4	0	65	1	8	21	0	0	3	2
	ang	0	79	2	1	1	1	16	0	8	0	5	28	5	39	15	0
	fea	0	33	48	2	3	10	3	1	42	0	23	10	4	13	8	0
	neu	4	2	0	90	0	2	2	0	13	1	2	76	0	1	4	3
	joy	0	52	8	4	4	22	10	0	2	0	1	10	56	29	0	2
	sur	1	35	4	0	0	31	28	1	4	1	2	42	15	28	7	1
	dis	0	36	4	6	6	6	41	1	19	0	15	39	0	9	17	1

Tabella 31: Matrice di confusione riportante le risposte agli stimoli suddivisi per lingua e per encoder

Per rendere più agevole la lettura dei dati, e quindi delle risposte fornite per ciascuna emozione e per ciascun insieme di risposte riportate in Tabella 31, e per far emergere la presenza di eventuali confusioni tra specifiche emozioni, si farà ricorso ad un procedimento di analisi dei cluster di tipo gerarchico (d'ora in avanti HCA, *hierarchical cluster analysis*).

La HCA costituisce un tipo di analisi che, nel caso qui trattato, fornisce un raggruppamento (o anche *clustering*) delle emozioni presentate che tenga conto di eventuali associazioni o confusioni tra di esse. Come risultato o *output* della HCA viene prodotta una particolare rappresentazione grafica che prende il nome di dendrogramma³¹⁷: oltre a fornire un grafico di facile ed immediata lettura delle relazioni tra i dati di volta in volta analizzati in

³¹⁷ Il dendrogramma ha una classica struttura ad albero in cui i vari elementi vengono raggruppati in base alla distanza che intercorre tra loro: il dendrogramma viene costruito partendo dai *rami* che costituiscono i casi singoli e iniziali, per arrivare, a seguito di fusioni che avvengono ad uno *stadio* di raggruppamento successivo nelle fasi intermedie, ad un unico ramo finale detto *radice*, con un ordine di lettura da sinistra a destra.

ciascuna matrice di confusione, consente anche di mettere in evidenza maggiori o minori gradi di confusione tra due o più emozioni organizzate e raggruppate secondo una struttura o rappresentazione grafica di tipo gerarchico.³¹⁸

10.5.2.1 Decoding di emozioni vocali in lingua italiana

Il *decoding* di produzioni vocali emotive da parte di soggetti di madrelingua italiana ha prodotto, supportando l'ipotesi dell'esistenza di un *in-group advantage*, la più alta percentuale media di corretto riconoscimento delle emozioni presentate (complessivamente pari al 61,57%)³¹⁹ rispetto alle altre tre lingue.

Andando ad analizzare le singole risposte con riferimento alle emozioni presentate (*cf.* Tabella 32) emerge come i rispondenti abbiano avuto difficoltà, per entrambi gli *encoder*, nel riconoscere l'emozione di gioia che viene confusa, nella maggior parte dei casi, con l'emozione di sorpresa (rispettivamente 36% per gli stimoli relativi all'*attore* e 24% per quelli relativi al *non-attore*).

Stimolo presentato		Risposta															
		Encoder = Attore								Encoder = Naif							
Lingua	Emozione	sad	ang	fea	neu	joy	sur	dis	n/a	sad	ang	fea	neu	joy	sur	dis	n/a
Italiano	sad	91	0	4	1	0	4	0	0	74	0	0	16	0	4	6	0
	ang	0	97	1	0	0	0	1	1	3	67	0	6	0	8	15	1
	fea	14	2	68	1	0	11	4	0	38	1	41	3	2	11	4	0
	neu	7	3	0	73	0	5	12	0	7	4	1	82	0	4	1	1
	joy	10	7	5	2	29	36	11	0	2	1	0	0	73	24	0	0
	sur	14	3	2	12	0	56	13	0	11	6	2	18	0	52	10	1
	dis	9	27	2	1	0	14	47	0	37	4	3	17	4	23	12	0

Tabella 32: Matrice di confusione delle risposte agli stimoli prodotti in lingua italiana da un attore e da un non-attore³²⁰

Riguardo le produzioni codificate dall'*attore* si rileva un alto grado di confusione dell'emozione di disgusto con la rabbia (27%). Per quelle codificate dal *non-attore* si registra, per la stessa emozione, un elevatissimo grado di confusione che porta gli ascoltatori a

³¹⁸ Per il processo di *clustering* è stata utilizzata la funzione HIERARCHICAL CLUSTER ANALYSIS del software statistico SPSS. Nello specifico sono state considerate come variabili le singole risposte fornite dai rispondenti raggruppate per caso (emozione codificata). I valori per ciascun caso, dopo essere stati trasformati e standardizzati in *z-scores* (con una media 0 e una deviazione standard 1), sono stati raggruppati utilizzando il metodo *Furthest neighbor* disponibile all'interno della funzione di SPSS.

³¹⁹ Si rammenta che il dato medio qui riportato risente della differenza di riconoscimento legata alla tipologia di *encoder* utilizzato nella produzione di emozioni vocali (*cf.* Tabella 30 e Figura 50).

³²⁰ Per questa tabella e per le altre che seguono, per evidenziare maggiormente i dati oggetto di discussione si è preferito colorare di rosso le celle in cui due emozioni (indifferentemente dall'*encoder*) sono state confuse dai soggetti ascoltatori con un valore percentuale uguale o maggiore a 20, mentre sono stati colorati di verde quei valori uguali o inferiori a 3 rilevati per entrambi gli *encoder*, ed assunti come indice di non-confusione tra due emozioni.

riconoscere il disgusto nel 37% dei casi con la tristezza e nel 23% dei casi con la sorpresa, facendo scendere la percentuale di corretto riconoscimento ad un bassissimo 12%. Anche per l'emozione di paura si rileva un'elevata confusione che la vede riconoscere nel 38% dei casi con l'emozione di tristezza.

Sempre con riferimento ai dati riportati in Tabella 32, è interessante notare come alcune emozioni non vengano mai confuse con altre. Per l'*encoder attore* la tristezza non viene mai confusa con rabbia, gioia, disgusto, ma solo con paura (4%), sorpresa (4%) e neutra (1%); la rabbia è in assoluto la più riconosciuta (97%) e raramente si confonde con le altre emozioni; la paura non viene mai confusa con la gioia e solo in pochissimi casi viene confusa con rabbia (2%), disgusto (4%) e la produzione neutra (1%); la produzione neutra non viene mai confusa con l'emozione di paura e gioia; la gioia viene raramente scambiata per una produzione neutra (2%); la sorpresa non viene mai scambiata con la gioia, mentre come già detto, si verifica il contrario (gioia viene confusa con sorpresa 36 volte su 100); per finire col disgusto che non si confonde mai con l'emozione di gioia e rarissimamente con l'emozione di paura (2%) o la produzione neutra (1%). Analoghe considerazioni, seppur con qualche sottile variazione, si possono avanzare per gli stimoli prodotti dall'*encoder naif*, evidenziando la presenza di una certa sistematicità non solo sulle confusioni che si manifestano per determinate emozioni, ma ancor più marcatamente per quelle che potrebbero essere definite non-confusioni.

Sottoponendo le due matrici di confusione in Tabella 32 ad una HCA (*cf.* Figura 51) si ha maggiore contezza della significatività e del grado di confusione tra le diverse emozioni che nella precedente esposizione possono risultare poco evidenti senza una indagine più approfondita e dettagliata.

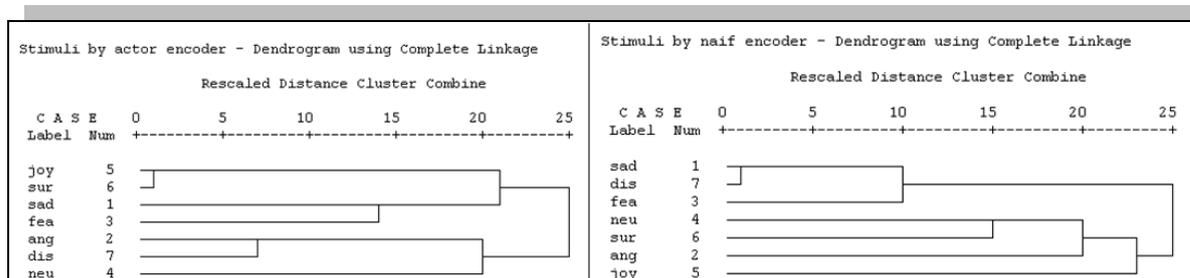


Figura 51: Clustering delle risposte agli stimoli vocali emotivi in lingua italiana prodotti rispettivamente da un attore e da un non-attore

Come appena detto, nella maggior parte dei casi i soggetti ascoltatori confondono, per entrambi gli *encoder*, l'emozione di gioia con l'emozione di sorpresa (36% e 24%). Tuttavia, al di là delle mere percentuali comunque leggibili in Tabella 32, questo dato assume un determinato "peso" in relazione alla percentuale di corretto riconoscimento e alla confusione

dell'emozione data con le altre emozioni. Infatti, se si osservano i due dendrogrammi riportati in Figura 51, emerge come le confusioni maggiori riguardano, per gli stimoli prodotti dall'*attore*, le emozioni di gioia e sorpresa (si veda il primo raggruppamento, o *ramo*, a sinistra più prossimo allo zero), seguiti da rabbia e disgusto (secondo *ramo*), tristezza e paura (terzo ramo); per il soggetto *non-attore* si ha, invece, un primo raggruppamento tra tristezza e disgusto (primo *ramo*), raggruppate a loro volta con paura e così via.

Con riferimento al dendrogramma relativo all'*attore*, gioia e sorpresa non vengono soltanto raggruppate perché maggiormente confuse, in questo caso, tra di loro (gioia viene confusa con sorpresa 36 volte su 100), ma anche e soprattutto perché vengono molto confuse con le altre emozioni: si vedano ad esempio, i valori di corretto riconoscimento per le due emozioni di gioia e sorpresa che nel complesso sono le più basse, rispettivamente del 29% e del 56%. Il raggruppamento successivo vede l'unione tra rabbia e disgusto: l'emozione di disgusto viene confusa con rabbia 27 volte su 100.

L'analisi delle matrici di confusione in Tabella 32 con una HCA, e la rappresentazione di tale analisi con l'ausilio dei dendrogrammi come in Figura 51, mette in evidenza il livello di confusione di determinate emozioni anche in rapporto alle altre: in tal senso, non ha solo rilevanza il fatto che una emozione "x" venga confusa con una emozione "y", ma assume notevole importanza anche quanto l'emozione "x" venga confusa con le altre emozioni "z_n". Si assiste, in poche parole, ad un ordine di importanza di tipo gerarchico in cui il grado di confusione tra due o più emozioni decresce all'allontanamento del raggruppamento (o *stadio*) dall'origine dei rami a sinistra (coincidente con lo zero) verso la *radice* del dendrogramma a destra. Ne consegue che le emozioni che vengono raggruppate ad un livello più alto (verso destra nel dendrogramma) hanno percentuali di corretto riconoscimento più elevate rispetto a quelle che, invece, si trovano raggruppate più a sinistra.

10.5.2.2 Decoding di emozioni vocali in lingua francese

Con riferimento alla Tabella 33 in cui vengono riportati per comodità i soli risultati per il francese (divisi per tipologia di *encoder*), le percentuali più alte di corretto riconoscimento si hanno per l'emozione di tristezza (rispettivamente 80% e 83%) e per la produzione neutra (rispettivamente 86% e 74%), mentre le più basse si hanno per l'emozione di rabbia (22% per l'*attore* e 17% per il *non-attore*) con un una sistematica confusione con l'emozione di sorpresa a cui si aggiunge anche la confusione con l'emozione di gioia per l'*attore* (34%) e l'emozione di disgusto per il soggetto *naif* (21%).

Inoltre, sia che si tratti di *attore*, sia che si tratti di soggetto *naif*, emerge come vi sia una certa sistematicità nel confondere:

- l'emozione di rabbia con l'emozione di sorpresa (rispettivamente 37% per l'*attore* e 36% per il soggetto *naif*), senza che si verifichi mai il contrario;
- l'emozione di sorpresa con la produzione neutra (rispettivamente 28% per l'*attore* e 23% per il soggetto *naif*), sebbene anche qui, a parte qualche isolato caso, non si verifica il contrario;
- l'emozione di disgusto con la produzione neutra (rispettivamente 30% per l'*attore* e 21% per il soggetto *naif*), mentre il contrario si verifica in un solo caso per entrambi gli *encoder*;
- l'emozione di disgusto con l'emozione di sorpresa (rispettivamente 18% per l'*attore* e 62% per il soggetto *naif*), dove anche in questo caso si verifica il contrario per pochissimi casi isolati.

Stimolo presentato		Risposta															
		Encoder = Attore								Encoder = Naif							
Lingua	Emozione	sad	ang	fea	neu	joy	sur	dis	n/a	sad	ang	fea	neu	joy	sur	dis	n/a
Francese	sad	80	1	9	4	1	3	2	0	83	0	4	2	0	4	4	3
	ang	0	22	0	1	34	37	4	2	2	17	0	15	7	36	21	2
	fea	1	1	86	3	0	5	2	2	14	1	1	47	1	31	4	1
	neu	9	0	0	86	1	3	1	0	18	0	2	74	0	2	1	3
	joy	2	0	3	0	72	23	0	0	0	1	0	7	29	63	0	0
	sur	5	0	1	28	16	47	3	0	4	0	1	23	4	64	3	1
	dis	21	1	0	30	5	18	23	2	6	1	0	21	8	62	1	1

Tabella 33: Matrice di confusione delle risposte agli stimoli prodotti in lingua francese da un attore e da un non-attore

I dati appena presentati per il francese trovano pienamente spiegazione nei due dendrogrammi in Figura 52.

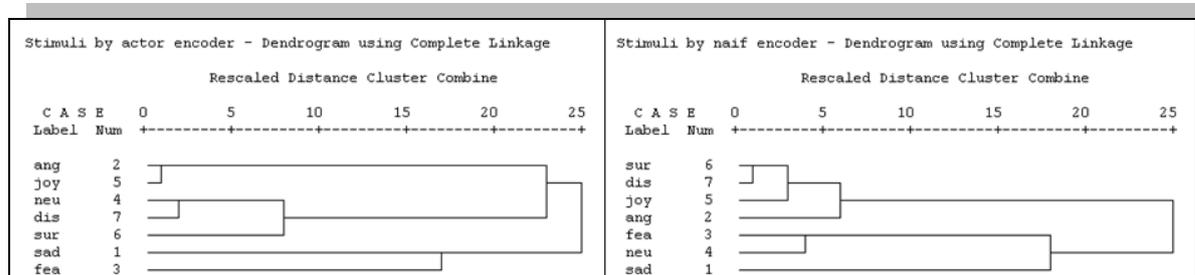


Figura 52: Clustering delle risposte agli stimoli vocali emotivi in lingua francese prodotti rispettivamente da un attore e da un non-attore

Si rilevano, infatti, una serie di raggruppamenti che si originano a *stadi* prossimi allo zero per entrambi gli *encoder* (maggiormente nel caso del soggetto *naif*) che, come detto, sono indice della presenza di un certo livello di confusione tra le emozioni raggruppate.

Oltre alle sistematicità sopra menzionate, per l'emozione di paura prodotta dal soggetto *naif* si registra un valore di corretto riconoscimento pressoché nullo (1%) se rapportato alla stessa emozione prodotta, invece, dal soggetto *attore* (86%), analogamente allo stimolo per l'emozione di disgusto che viene riconosciuto in solo caso su 100 (1%).

Come già evidenziato per gli stimoli in lingua italiana, anche per gli stimoli in lingua francese si assiste ad una non-confusione di specifiche emozioni con altre. Per entrambi gli *encoder* e con lievissime oscillazioni dell'ordine di non più di 3 punti percentuali si può notare come certe emozioni non vengano mai confuse con altre:

- la tristezza non viene mai confusa con rabbia e gioia;
- la rabbia non si confonde con l'emozione di tristezza e la produzione neutra;
- la paura non viene mai confusa con la rabbia e la gioia;
- la produzione neutra con nessuna delle emozioni presentate, tranne che con l'emozione di tristezza;
- l'emozione di gioia viene nettamente distinta da tristezza, rabbia, paura e disgusto;
- la sorpresa non viene mai riconosciuta con la rabbia, la paura e il disgusto;
- il disgusto non viene confuso con rabbia e paura.

10.5.2.3 Decoding di emozioni vocali in lingua inglese

Passando agli stimoli presentati in lingua inglese (*cf.* Tabella 34) si rileva una certa sistematicità nel confondere:

- la produzione neutra con quella di tristezza (rispettivamente 41% per l'*attore* e 30% per il *non-attore*);
- l'emozione di gioia con la sorpresa (rispettivamente 61% e 54%);
- e ancora l'emozione di disgusto con la rabbia (rispettivamente 46% 41%).

Per il primo e il terzo caso appena presentati emerge anche come si verifichi l'esatto contrario, ovvero di confusione dell'emozione di tristezza con la produzione neutra (10% per l'*attore* e 25% per il soggetto *naif*, e di rabbia con l'emozione di disgusto (rispettivamente 11% per l'*attore* e 22% per il *non-attore*), mentre per il secondo in elenco si registrano solo pochissimi casi di confusione, per entrambi gli *encoder*, di sorpresa con gioia (2% e 4%).

Come era già stato rilevato per le risposte all'emozione di paura prodotta dal *non-attore* francese, anche per il *non-attore* inglese si registra un valore di corretto riconoscimento pressoché nullo (4%) che in questo caso viene per lo più confusa con la paura (33%) e la

produzione neutra (40%), mentre raggiunge la stessa emozione prodotta dal soggetto *attore* raggiunge il 71% di risposte corrette.

Stimolo presentato		Risposta															
		Encoder = Attore								Encoder = Naif							
Lingua	Emozione	sad	ang	fea	neu	joy	sur	dis	n/a	sad	ang	fea	neu	joy	sur	dis	n/a
Inglese	sad	55	1	17	10	0	7	10	0	44	3	4	25	0	9	13	2
	ang	0	78	6	0	1	3	11	1	0	40	0	11	0	25	22	2
	fea	10	0	71	0	0	15	2	2	33	1	4	40	0	11	9	2
	neu	41	1	7	38	0	8	5	0	30	0	3	62	0	0	4	1
	joy	2	11	4	0	8	61	14	0	0	2	0	3	30	54	11	0
	sur	5	18	16	2	2	39	17	1	4	5	2	26	4	40	17	2
	dis	1	46	0	1	1	7	44	0	0	41	1	2	1	21	34	0

Tabella 34: Matrice di confusione delle risposte agli stimoli prodotti in lingua inglese da un attore e da un non-attore

In aggiunta a quanto già esposto, per gli stimoli del soggetto *non-attore* si rilevano una serie di ulteriori confusioni che vedono l'emozione di rabbia confusa con l'emozione di sorpresa (25%), l'emozione di sorpresa confusa con la produzione neutra (26%), l'emozione di disgusto con l'emozione di sorpresa (21%).

Per entrambi gli *encoder* si nota (celle in verde) determinate emozioni, anche in questo caso non vengano mai (o quasi mai) confuse con altre:

- l'emozione di tristezza non viene mai confusa con rabbia e gioia;
- l'emozione di rabbia non si confonde con tristezza e gioia;
- la paura viene sempre distinta da rabbia e gioia;
- la produzione neutra si distingue nettamente da rabbia e gioia;
- l'emozione di gioia non viene confusa con tristezza e la produzione neutra;
- il disgusto viene distinto da tutte le emozioni tranne che per la rabbia e per la sorpresa.

Sulla matrice di confusione in Tabella 34 è stata quindi operata una HCA.

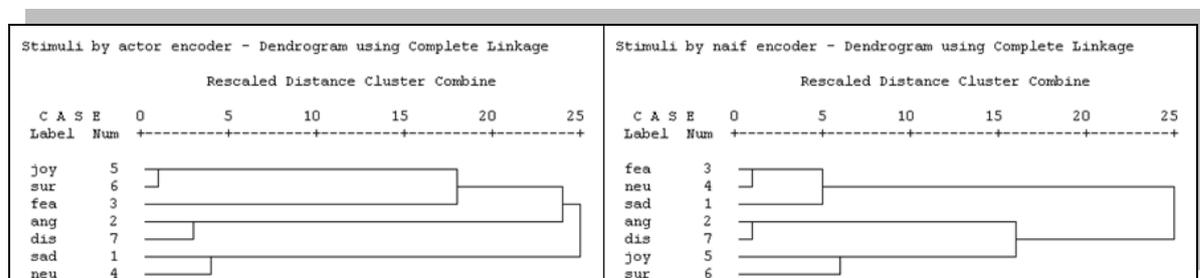


Figura 53: Clustering delle risposte agli stimoli vocali emotivi in lingua inglese prodotti rispettivamente da un attore e da un non-attore

Dai risultati della HCA, di cui viene proposto il dendrogramma in Figura 53, si evince chiaramente come per le risposte agli stimoli prodotti dall'*attore* inglese vi sia un primo

raggruppamento tra gioia e sorpresa: le risposte di corretto riconoscimento per l'emozione di gioia sono, infatti, pari all'8% con un'elevata percentuale di confusione di quest'ultima con l'emozione di sorpresa (61%). Seguono poi rabbia e disgusto (dove la seconda viene confusa con la prima con una percentuale del 46%) e l'emozione di tristezza con la produzione neutra (in cui la seconda viene confusa al 41% con la prima). In modo del tutto analogo si registrano, per gli stimoli prodotti dal soggetto *naif*, due raggruppamenti iniziali che si originano ad uno stesso stadio, ovvero tra l'emozione di paura e la produzione neutra e tra le emozioni di rabbia e di disgusto a cui seguono tutte le altre.

10.5.2.4 Decoding di emozioni vocali in lingua tedesca

Le percentuali di corretto riconoscimento relative agli stimoli prodotti in lingua tedesca (*cf.* Tabella 35) hanno fatto registrare per entrambi gli *encoder*, valori superiori al caso tranne che per l'emozione di gioia prodotta dal soggetto *attore* (4%) e per l'emozione di rabbia prodotta dal *non-attore* (0%). Nel primo caso si registra, infatti, una elevatissima confusione dell'emozione di gioia con l'emozione di rabbia (52%) e con l'emozione di sorpresa (22%), anche se non si verifica il contrario, ovvero di confondere rabbia o sorpresa con gioia. Nel secondo caso, invece, l'emozione di rabbia viene per lo più confusa con la produzione neutra (28%) e l'emozione di sorpresa (39%), ma nemmeno in tal caso si verifica mai il contrario (tranne che in due casi molto isolati).

Stimolo presentato		Risposta															
		Encoder = Attore								Encoder = Naif							
Lingua	Emozione	sad	ang	fea	neu	joy	sur	dis	n/a	sad	ang	fea	neu	joy	sur	dis	n/a
Tedesco	sad	67	2	18	7	0	2	4	0	65	1	8	21	0	0	3	2
	ang	0	79	2	1	1	1	16	0	8	0	5	28	5	39	15	0
	fea	0	33	48	2	3	10	3	1	42	0	23	10	4	13	8	0
	neu	4	2	0	90	0	2	2	0	13	1	2	76	0	1	4	3
	joy	0	52	8	4	4	22	10	0	2	0	1	10	56	29	0	2
	sur	1	35	4	0	0	31	28	1	4	1	2	42	15	28	7	1
	dis	0	36	4	6	6	6	41	1	19	0	15	39	0	9	17	1

Tabella 35: Matrice di confusione delle risposte agli stimoli prodotti in lingua tedesca da un attore e da un non-attore

Per gli stimoli in lingua tedesca viene meno quella sistematicità che vedeva, per entrambi gli *encoder*, un certo grado di confusione tra specifiche emozioni. L'unico errore di riconoscimento sistematico che viene qui identificato riguarda l'emozione di gioia che viene confusa, sia per l'uno che per l'altro *encoder*, rispettivamente nel 22% e nel 29% dei casi, con l'emozione di sorpresa.

Osservando i dati in Tabella 35 emerge quello che nell'immaginario comune dell'italiano medio vede il tedesco come quasi sempre arrabbiato: nella maggior parte dei casi, per gli stimoli prodotti dall'attore, si assiste ad una elevata confusione di quasi tutte le emozioni presentate con l'emozione di rabbia, ad eccezione delle emozioni di tristezza e della produzione neutra. Il dato rilevato per le risposte agli stimoli dell'attore non trova però supporto nel soggetto *naif* dove la stessa emozione di rabbia non viene mai riconosciuta con l'emozione intesa e dove persino anche le altre emozioni non vengono mai confuse con essa. In quest'ultimo caso, per il soggetto *non-attore* si rilevano elevate confusioni delle emozioni di tristezza, rabbia, sorpresa e disgusto con la produzione neutra (rispettivamente con le percentuali del 21%, 28%, 42% e 39%). Sempre per il soggetto *naif* si assiste ad una significativa confusione dell'emozione di paura con l'emozione di tristezza (42%).

Per gli stimoli in lingua tedesca viene anche meno la sistematicità riscontrata riguardo quelle che sono state definite non-confusioni, che in questo caso si riduce sostanzialmente per gli entrambi gli encoder a tre casi:

- l'emozione di tristezza non si confonde con le emozioni di rabbia, gioia e sorpresa;
- la produzione neutra non viene quasi mai confusa con le altre emozioni, tranne che con tristezza (4% per l'attore e 13% per il non-attore) e disgusto (rispettivamente 2% e 4%);
- l'emozione di gioia, infine, non si confonde con l'emozione di tristezza.

Anche per gli stimoli analizzati in quest'ultimo caso per la lingua tedesca, i dati dell'HCA supportano le analisi e le considerazioni fin qui fatte.

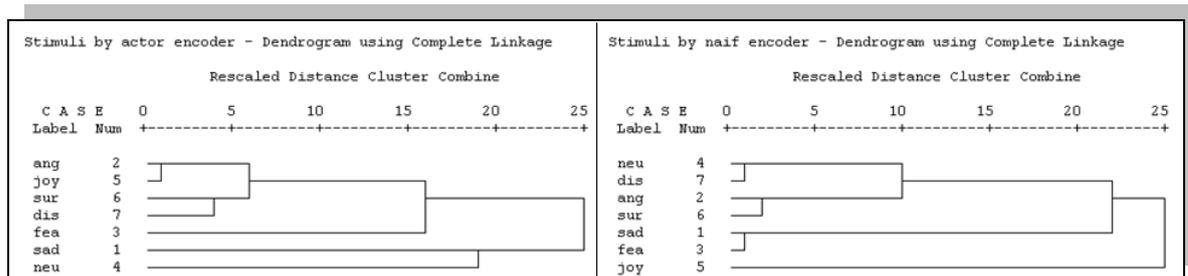


Figura 54: Clustering delle risposte agli stimoli vocali emotivi in lingua tedesca prodotti rispettivamente da un attore e da un non-attore

Per l'*attore* si rileva, infatti, un ampio raggruppamento di emozioni con l'emozione di rabbia evidenziando quanto più sopra detto riguardo l'elevata confusione che con essa si ha, in ordine decrescente per le emozioni di gioia, disgusto, sorpresa e paura.

La medesima osservazione può essere fatta per il soggetto naif dove è la produzione neutra l'emozione con cui vengono identificate le emozioni di sorpresa, disgusto, rabbia e tristezza.

10.5.3 La competenza dei rispondenti nelle lingue straniere presentate

Per verificare se la maggiore o minore abilità dei soggetti ascoltatori a decodificare correttamente gli stimoli emotivi presentati nelle tre lingue straniere qui prese in considerazione fosse in qualche modo riconducibile alla loro competenza in ciascuna delle lingue, ai rispondenti è stato chiesto di indicare, appunto, il loro grado di competenza: per tutte e tre le lingue le opzioni possibili erano complessivamente 6 (*cf.* § 10.2).

Tavola di contingenza Competenza * Lingua				
<i>Competenza</i>	English	French	German	Totale
Nessuna	1	29	68	98
Apprendista	0	13	7	20
Scolastica	18	35	9	62
Universitaria	44	14	12	70
Avanzata	33	8	2	43
Madrelingua	4	1	2	7
<i>Totale</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>300</i>

Tabella 36: Tavola di contingenza per la competenza dei rispondenti nelle tre lingue straniere presentate

Esaminando la tavola di contingenza relativa alla competenza dei rispondenti in ciascuna delle tre lingue in Tabella 36 emerge come:

- la maggior parte di essi non ha la minima competenza nella lingua tedesca (68%) a cui segue la lingua francese (29%);
- una buona fetta di essi ha una competenza di base nella lingua francese con un 35% di competenze scolastiche;
- un'altra buona percentuale ha, invece, una competenza medio-alta nella lingua inglese (con un 44% di competenza universitaria e un 33% di competenza avanzata).

I dati appena presentati, di fatto, sembrano rappresentare benissimo la situazione italiana in riferimento alla conoscenza o alle competenze che si ha nelle lingue straniere. Sebbene non sia questo l'obiettivo della ricerca, dai dati raccolti viene fuori uno spaccato assai eloquente del sistema istruzione italiano che vede un netto predominio dell'insegnamento dell'inglese ai livelli più alti dei gradi di istruzione e oltre (ricoprendo ben il 77% del presente campione di indagine), una prevalenza del francese al livello scolastico coincidente con le scuole dell'obbligo (35%). Quanto detto si evince ancor di più riportando i

risultati appena presentati su di un grafico (cfr. Figura 55) che consente di cogliere visivamente tali differenze.

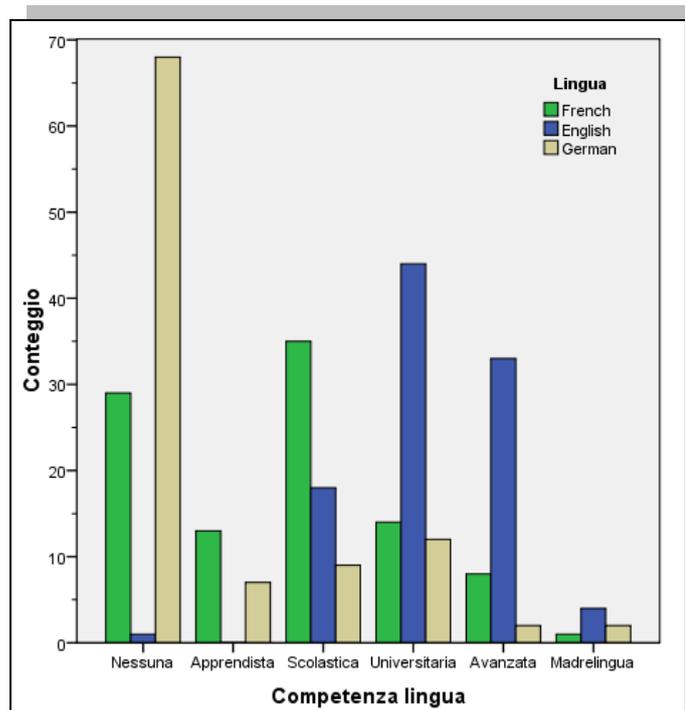


Figura 55: Competenza dei soggetti rispondenti nelle tre lingue straniere presentate

Da una prima analisi dei dati, sebbene si rendano necessari ulteriori approfondimenti, sembra essere contraddetta la possibilità che ad una competenza maggiore in una lingua corrisponda una maggiore abilità nel riconoscere correttamente gli stimoli emotivi ad essa appartenenti. Confrontando i dati di corretto riconoscimento in Tabella 30 più sopra riportata emerge, infatti, come ad un livello di competenza maggiore nella lingua inglese (77% considerando sia la competenza universitaria che avanzata) corrisponde una percentuale media di corretto riconoscimento degli stimoli nella medesima lingua del 41,93%, percentuale inferiore, quindi, a quella rilevata per gli stimoli in lingua tedesca (44,64%) a cui corrisponde una competenza pari a zero per il 68% dei rispondenti (75% se si includono anche i soggetti con competenza da apprendista).

Per verificare statisticamente l'ipotesi di indipendenza tra numero di risposte corrette e livello di competenza per ciascuna lingua, che dalle prime analisi sin qui condotte sembra essere confermata, si è dapprima proceduto al computo di una tavola di contingenza dei casi osservati: in intestazione di colonna è stato riportato il numero totale di risposte corrette registrate (da 20 a 42, su un massimo possibile di 56) e in intestazione di riga le lingue (francese, inglese e tedesco) raggruppate, a loro volta, per competenza (nessuna, apprendista, scolastica, universitaria, avanzata, madrelingua) come riportato di seguito in Tabella 37.

Tavola di contingenza Lingua * Total score * Competenza																					
Competenza	Lingua	Total score																			Totale
		20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	42	
Nessuna	English	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
	French	2	2	1	2	3	2	3	3	3	0	1	2	1	2	1	1	0	0	0	29
	German	3	4	4	6	4	5	4	8	7	7	5	2	1	2	0	1	3	1	1	68
	<i>Totale</i>	5	6	5	8	7	7	7	12	10	7	6	4	2	4	1	2	3	1	1	98
Apprendista	French	1	0		1	1	3	0	1	1	2			0		1	0	2			13
	German	0	1		0	0	0	1	2	0	0			1		1	1	0			7
	<i>Totale</i>	1	1		1	1	3	1	3	1	2			1		2	1	2			20
Scolastica	English		1	2	1	0	2	1	1	1	3	0	1	1	1	0	2	1		0	18
	French		1	4	2	1	0	4	4	4	4	2	3	2	0	1	1	1		1	35
	German		1	1	0	0	1	0	0	0	2	0	1	1	0	2	0	0		0	9
	<i>Totale</i>		3	7	3	1	3	5	5	5	9	2	5	4	1	3	3	2		1	62
Universitaria	English	2	5	0	2	2	3	3	6	7	2	3	5	2	1	0	0		1		44
	French	0	2	1	1	0	1	0	2	1	1	2	1	1	0	0	0		1		14
	German	0	0	0	0	2	0	2	1	2	0	1	2	0	0	1	1		0		12
	<i>Totale</i>	2	7	1	3	4	4	5	9	10	3	6	8	3	1	1	1		2		70
Avanzata	English	1	0	3	3	3	1	2	3	1	5	3		1		3	1	2		1	33
	French	0	1	0	0	1	0	0	1	0	2	1		0		1	1	0		0	8
	German	0	0	1	0	0	0	0	0	0	1	0		0		0	0	0		0	2
	<i>Totale</i>	1	1	4	3	4	1	2	4	1	8	4		1		4	2	2		1	43
Madrelingua	English			1		1		1			0		0	0		1					4
	French			0		0		0			1		0	0		0					1
	German			0		0		0			0		1	1		0					2
	<i>Totale</i>			1		1		1			1		1	1		1					7

Tabella 37: Tavola di contingenza dei casi osservati per il numero totale di risposte corrette in funzione della competenza dei rispondenti in ciascuna delle tre lingue straniere presentate

Sui dati calcolati in Tabella 37 è stato quindi possibile applicare un test chi-quadro di indipendenza per verificare l'ipotesi H0 di indipendenza tra le variabili prese in considerazione ($\alpha = 0,05$).

Chi-quadrato				
Competenza		Valore	df	Sig. asint. (2 vie)
Nessuna	Chi-quadrato di Pearson	19,406 ^a	36	,989
	Rapporto di verosimiglianza	19,596	36	,988
	N. di casi validi	98		
Apprendista	Chi-quadrato di Pearson	14,872 ^b	12	,249
	Rapporto di verosimiglianza	19,306	12	,081
	N. di casi validi	20		
Scolastica	Chi-quadrato di Pearson	25,870 ^c	32	,769
	Rapporto di verosimiglianza	29,339	32	,602
	N. di casi validi	62		
Universitaria	Chi-quadrato di Pearson	29,490 ^d	32	,594
	Rapporto di verosimiglianza	31,016	32	,516
	N. di casi validi	70		
Avanzata	Chi-quadrato di Pearson	16,206 ^e	30	,981
	Rapporto di verosimiglianza	16,980	30	,973
	N. di casi validi	43		
Madrelingua	Chi-quadrato di Pearson	14,000 ^f	12	,301
	Rapporto di verosimiglianza	13,380	12	,342
	N. di casi validi	7		

a. 54 celle (94,7%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è ,01.
b. 26 celle (100,0%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è ,35.
c. 50 celle (98,0%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è ,15.
d. 48 celle (94,1%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è ,17.
e. 47 celle (97,9%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è ,05.
f. 21 celle (100,0%) hanno un conteggio atteso inferiore a 5. Il conteggio atteso minimo è ,14.

Tabella 38: Risultati della statistica chi-quadro per i diversi livelli di competenza dei rispondenti nelle lingue straniere presentate

Dai risultati riportati in Tabella 38 per il test chi-quadro di indipendenza si rilevano, per tutti i livelli di competenza considerati, valori di significatività superiori alla soglia prefissata ($\alpha = 0,05$), dato che porta ad accettare l'ipotesi nulla H0 che vi sia indipendenza tra le variabili esaminate.

Pertanto, si può affermare che tra numero di risposte complessivamente fornite e livello di competenza nelle lingue straniere presentate vi sia assoluta indipendenza.

10.5.4 Alla ricerca di differenze legate al sesso dei rispondenti (*gender-effects*)

Nella formulazione delle ipotesi per questa fase della ricerca (vedi § 8.2), ci si era anche interrogati se vi fossero differenze nell'abilità di decodificare produzioni vocali emotive riconducibili al sesso dei soggetti ascoltatori (*gender-effects*).

Il grafico a dispersione riportato di seguito in Figura 56 evidenzia un certo grado di omogeneità nel numero totale di risposte corrette fornite escludendo, a prima vista, possibili influenze riconducibili al sesso.

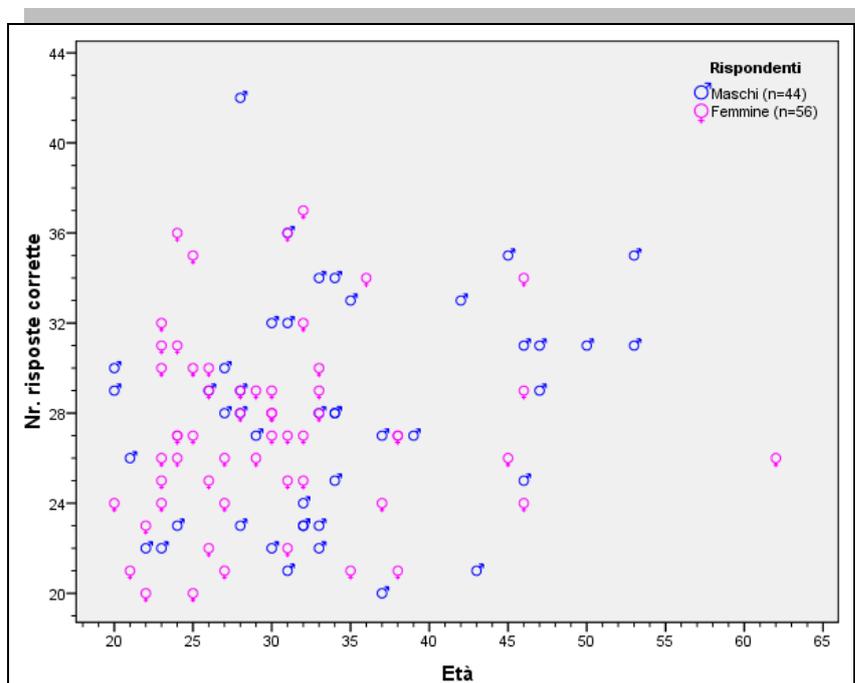


Figura 56: Grafico a dispersione del numero di risposte corrette fornite dai rispondenti (divisi per sesso) in funzione dell'età

Le statistiche di gruppo riportate in Tabella 39 evidenziano come mediamente il numero totale di risposte corrette (*Total score*)³²¹ sia effettivamente assai simile per entrambi i gruppi (Femmine e Maschi).

Statistiche di gruppo					
	Sesso_group	N	Media	Deviazione std.	Errore std. Media
Total score	Femmina	56	27,27	4,136	,553
	Maschio	44	28,00	4,899	,739

Tabella 39: Statistiche di gruppo per il numero totale di risposte corrette fornite dai maschi e dalle femmine

Tuttavia, per verificare se la differenza di sesso potesse essere comunque ritenuta significativamente rilevante in riferimento al numero totale di risposte correttamente fornite si è proceduto ad un calcolo del *Test-T per campioni indipendenti* (valore di $\alpha = 0,05$).

³²¹ Si rammenta il punteggio massimo di 56 nel caso decodifica corretta di tutti gli stimoli presentati (56 in tutto).

Test per campioni indipendenti										
		Test di Levene di uguaglianza delle varianze		Test t di uguaglianza delle medie						
		F	Sig.	t	df	Sig. (2-code)	Diff. fra medie	Diff. err. st.	Intervallo di confidenza per la differenza al 95%	
									Inf.	Sup.
Total score	Assumi varianze uguali	1,674	,199	-,810	98	,420	-,732	,904	-2,526	1,062
	Non assumere varianze uguali			-,794	84,045	,430	-,732	,922	-2,567	1,102

Tabella 40: Risultati del Test-T per campioni indipendenti per il numero totale di risposte corrette fornite dai due gruppi di rispondenti (Femmine e Maschi)

I risultati del test, riportati in Tabella 40, confermano quanto appena detto riguardo il grado di omogeneità del numero totale di risposte corrette fornite dai due gruppi (vedi *Test di Levene di uguaglianza delle varianze*: $F= 1,674$; $\alpha = 0,199$). In conclusione a questo aspetto si può, quindi, affermare, che il risultato del Test-T (*Sig. a due code* = 0,420) conferma l'ipotesi, sopra richiamata, che non vi siano differenze di abilità nella decodifica di espressioni vocali emotive riconducibili al sesso: si esclude, pertanto, la presenza di *gender-effects* sul numero di risposte corrette complessivamente fornite al test percettivo somministrato.

10.5.5 Impressioni soggettive da parte dei rispondenti

Prendendo spunto da Pell *et al.* (2009), si è pensato di verificare se vi fosse relazione tra i dati del test ed eventuali impressioni soggettive dei rispondenti in merito all'esperimento svolto. Per questa ragione, nella terza fase del test, ai rispondenti sono state poste delle domande (*cfr.* § 10.2) per verificare:

1. quale lingua fosse stata per loro la più difficile da riconoscere;
2. quali fossero le due emozioni ritenute più difficili da riconoscere a prescindere dalla lingua.

Di seguito l'analisi delle risposte alle due domande.

10.5.5.1 La lingua più difficile da riconoscere

Nel complesso, le risposte a questa domanda supportano solo in parte i dati riguardanti il corretto riconoscimento degli stimoli presentati. Da quanto si evince dai dati rappresentati nel grafico in Figura 57, si rileva come il 47% dei rispondenti abbia individuato nella lingua

tedesca la lingua più difficile da riconoscere, anche se la percentuale media di riconoscimento corretto per il tedesco contraddice in parte l'impressione dei soggetti ascoltatori: sono, infatti, gli stimoli in lingua inglese ad essere meno riconosciuti (anche se di poco con un 41,93%) rispetto a quelli in lingua tedesca (44,64%). Ciononostante, il dato che vede il tedesco come lingua più difficile da riconoscere è coerente con la più bassa competenza registrata (cfr. § 10.5.3) per la stessa lingua.

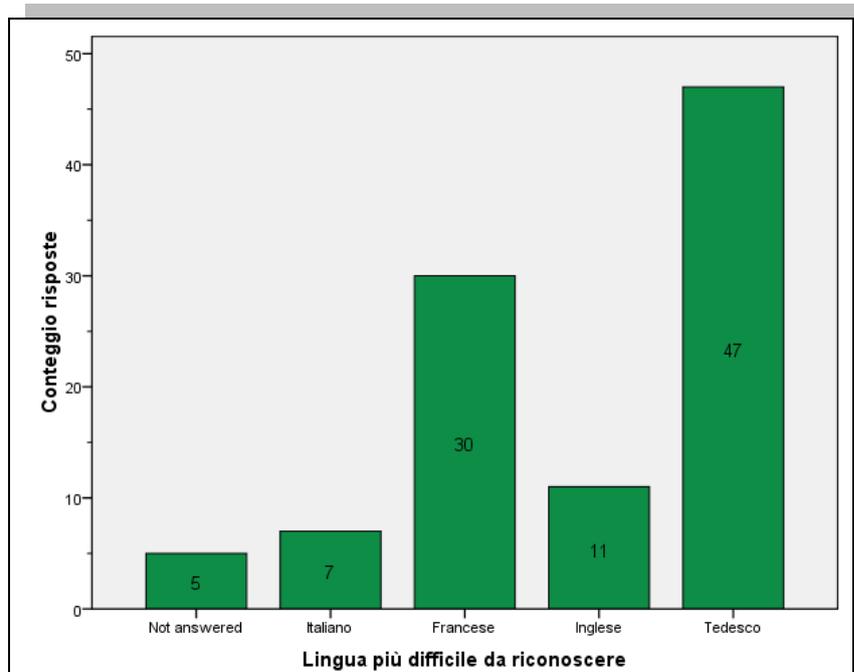


Figura 57: Conteggio delle risposte per la lingua ritenuta più difficile da riconoscere da parte dei soggetti ascoltatori

Alla lingua tedesca segue la lingua francese che viene considerata difficile da riconoscere dal 30% dei soggetti ascoltatori, mentre solo l'11% di essi considera l'inglese come lingua più difficile da riconoscere. Anche in questo caso i dati riguardanti il corretto riconoscimento degli stimoli emotivi nelle due lingue sembrano smentire questa sensazione espressa da parte dei rispondenti. Nonostante l'inglese venga considerato meno difficile da riconoscere del francese, e nonostante nella lingua inglese buona parte dei soggetti (77%) abbia una competenza medio-alta, si registra un corretto riconoscimento complessivamente inferiore (41,93%) in rapporto allo stesso dato per la lingua francese (48,93%), nella quale si registra una competenza medio-bassa (tra competenza scolastica e da apprendista si arriva ad un 48% dei rispondenti).

Infine, nonostante sia alquanto strano, risulta interessante evidenziare il fatto che qualcuno dei rispondenti (7%) abbia indicato nella lingua italiana la lingua più problematica. È probabile che i rispondenti avessero, nella situazione presentata, maggiori aspettative nei confronti degli stimoli appartenenti alla propria lingua, aspettativa che per le difficoltà

incontrate nello svolgimento del task è stata forse disattesa (si vedano ad esempio alcuni dei commenti dei rispondenti riportati Appendice X).

10.5.5.2 L'emozione più difficile da riconoscere

L'altra domanda posta ai rispondenti prevedeva che gli stessi indicassero due delle emozioni che, a prescindere dalla lingua, essi ritenessero più difficili da riconoscere.

Come già accennato, sebbene fosse richiesto di indicare soltanto due emozioni, dal computo dei dati riportati nella tavola di contingenza in Tabella 41 emerge come vi sia una certa disomogeneità riguardo al numero di risposte fornite: si va da una sola risposta ad un massimo di quattro.

Tavola di contingenza		
	Conteggio	Totale
Emozione	Disgusto	3
	Gioia	1
	Gioia:Disgusto	7
	Gioia:Sorpresa	7
	Gioia:Sorpresa:Disgusto	1
	Neutra (nessuna emozione)	3
	Neutra (nessuna emozione):Disgusto	7
	Neutra (nessuna emozione):Gioia	2
	Neutra (nessuna emozione):Gioia:Sorpresa	1
	Neutra (nessuna emozione):Sorpresa	1
	Paura:Disgusto	24
	Paura:Gioia	3
	Paura:Gioia:Disgusto	2
	Paura:Neutra (nessuna emozione)	1
	Paura:Neutra (nessuna emozione):Disgusto	3
	Paura:Sorpresa	2
	Paura:Sorpresa:Disgusto	1
	Rabbia:Disgusto	7
	Rabbia:Paura:Sorpresa:Disgusto	1
	Rabbia:Sorpresa	1
	Sorpresa:Disgusto	9
	Tristezza:Disgusto	4
	Tristezza:Neutra (nessuna emozione)	3
	Tristezza:Paura	2
	Tristezza:Rabbia	2
	Tristezza:Rabbia:Sorpresa:Disgusto	1
	Tristezza:Sorpresa:Disgusto	1
	Totale	100

Tabella 41: Tavola di contingenza per le risposte date dai soggetti ascoltatori alla domanda concernente l'emozione più difficile da riconoscere

Dalle risposte dei soggetti ascoltatori in Tabella 41 emerge come il maggior numero di risposte fornite riguardi le emozioni di paura e disgusto (24%) anche se le appena citate emozioni appaiono anche in altre risposte in associazione ad altre emozioni. Per questa

ragione si è proceduto ad un calcolo delle risposte per ciascuna emozione in termini assoluti, risultato che viene riportato in Tabella 42.

Emozione	Nr. risposte
Disgusto	71
Rabbia	12
Paura	39
Tristezza	13
Neutra (nessuna emozione)	21
Gioia	24
Sorpresa	26

Tabella 42: Riepilogo complessivo delle risposte alla domanda sulle emozioni più difficili da riconoscere

Analizzando nello specifico i dati Tabella 42 resta confermato il maggior numero di risposte per l'emozione di disgusto (71%) e di paura (39%), mentre le più basse percentuali riguardano le emozioni di rabbia (12%) e di tristezza (13%). Per un maggiore apprezzamento delle risposte fornite dai rispondenti alla domanda sull'emozione più difficile da riconoscere a prescindere dalla lingua in Figura 58 viene fornita una rappresentazione grafica dei dati riportati in Tabella 42.

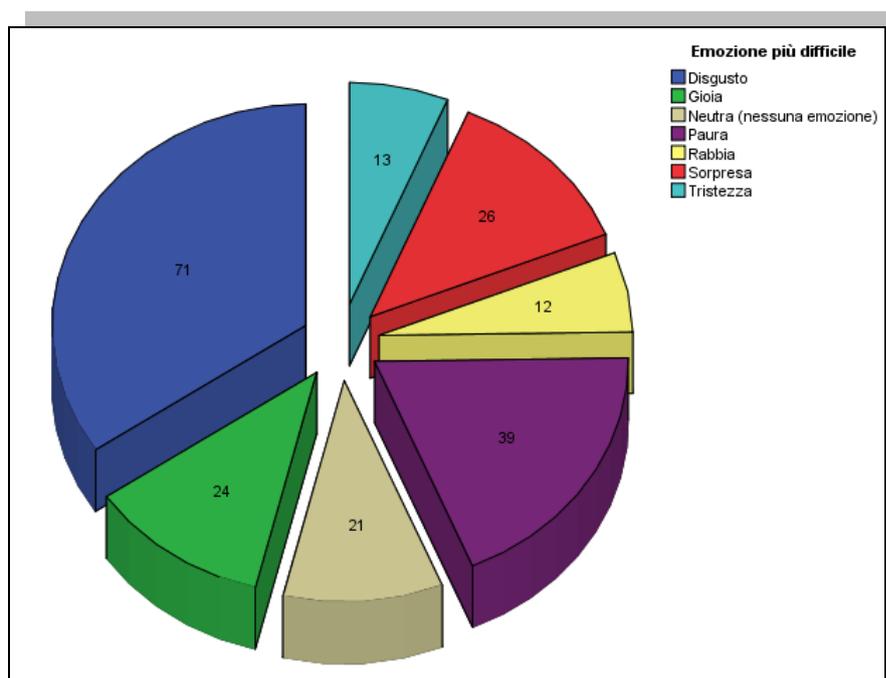


Figura 58: Grafico a torta relativo alle risposte complessive sulle emozioni più difficili da riconoscere

L'elevato numero di risposte per l'emozione di disgusto (71%) trova conferma nel basso numero di risposte corrette fornite in ciascuna lingua per la stessa emozione che raggiunge complessivamente il 27,5%³²². Di contro, al basso numero di risposte che identifica

³²² Con percentuali del 30% per l'italiano, del 12% per il francese, del 29% per l'inglese e del 39% per il tedesco.

le emozioni di rabbia e di tristezza come le meno difficili da riconoscere (rispettivamente 12% e 13%) corrispondono percentuali di riconoscimenti corretto complessivamente più alte (50,25% e 70,25%).

10.6 Conclusioni

I risultati sin qui ottenuti e presentati sono da ritenersi più che soddisfacenti. Nonostante l'elevato numero di emozioni utilizzate nella sperimentazione (complessivamente 7: *anger, disgust, fear, joy, neutral, sadness, surprise*) e la duplice tipologia di encoder (attori e non-attori) confermano pienamente i risultati di precedenti ricerche condotte nello stesso ambito.

Oltre ad aver documentato l'abilità dei soggetti a decodificare meglio gli stimoli emotivi prodotti nella propria lingua madre, rivelando la presenza di quello che in letteratura viene definito *in-group advantage*, è stata altresì documentata l'abilità degli esseri umani a decodificare stimoli emotivi anche in lingue diverse dalla propria con percentuali medie di corretto riconoscimento superiori a quelle attese dal caso.

Dalla somministrazione dell'esperimento qui presentato e dalle analisi condotte è stato escluso che il grado di conoscenza e competenza della lingua presentata possa in qualche modo influire sul riconoscimento delle produzioni emotive presentate, lasciando per il momento aperta la questione ipotizzata da Scherer *et al.* (2001) riguardo la possibile influenza nel riconoscimento riconducibile ad aspetti segmentali e soprasegmentali.

È stata inoltre verificata anche per il canale vocale l'ipotesi che, in precedenti ricerche, attribuiva maggiore abilità ai soggetti di sesso femminile nel decodificare emozioni espresse attraverso il volto: dalle analisi statistiche condotte questa ipotesi non trova supporto nel riconoscimento di stimoli emotivi vocali escludendo quindi la presenza di differenze di riconoscimento riconducibili al sesso dei soggetti ascoltatori (*gender-effects*).

Il fatto che per i soggetti *non-attori* si siano registrati valori di riconoscimento corretto nel complesso più bassi rispetto al quelli dell'*attore*, sembra confermare un assunto al quale più volte si è accennato (vedi Wilting *et al.*, 2006) che vedrebbe negli *attori* una indubbia maggiore abilità nel codificare o produrre emozioni prototipiche a vario titolo elicitate o indotte e quindi una maggiore riconoscibilità delle produzioni da questi prodotte.

In futuro sarà inoltre interessante verificare se i risultati rilevati in questa sede per gli stimoli presentati nelle quattro lingue a soggetti ascoltatori di madre lingua italiana siano ottenibili sottoponendo lo stesso esperimento a soggetti ascoltatori di madrelingua diversa

(francese, inglese e tedesco) a completamento dello studio intrapreso nella direzione proposta da Scherer *et al.* (2001: 88) con la costruzione di una “*encoder-decoder-emotion matrix*”.

11 Step 3 - Analisi dei correlati acustici delle produzioni emotive mistilingui

Dai risultati della somministrazione dell'esperimento di *decoding* cross-linguistico di cui al precedente capitolo, è emerso come i 100 rispondenti, tutti di madrelingua italiana, siano stati in grado di decodificare gli stimoli emotivi loro presentati nella propria lingua e in lingue diverse dall'italiano con percentuali complessivamente superiori a quelle che ci si sarebbe aspettati se gli stessi avessero risposto a caso. Inoltre, è stato dimostrato come, nel presente caso, l'abilità di riconoscere emozioni vocali in una lingua diversa dalla propria sia assolutamente indipendente dalle competenze dei soggetti ascoltatori nelle lingue esaminate.

Come era già stato parzialmente ipotizzato sulla base di precedenti ricerche condotte nella stessa direzione della presente proposta, e come appurato e testimoniato dai risultati scaturiti dall'attuazione del secondo *step* della ricerca con l'esperimento di *decoding*, è assai evidente che la capacità dell'uomo nel riconoscere emozioni trasmesse attraverso la sola voce sia in buona parte da ricondurre alla presenza di specifici correlati acustici. Quanto appena detto risulta ulteriormente confermato dalla capacità dei soggetti ascoltatori nel decodificare correttamente anche emozioni vocali in una lingua diversa dalla propria con differenze riconducibili principalmente alla lingua, all'emozione e all'encoder presentato.

Date queste premesse, avendo escluso la possibilità che l'abilità di riconoscere gli stimoli vocali emotivi presentati fosse in un certo qual modo dovuta o riconducibile al livello di competenza nelle singole lingue da parte dei rispondenti, si è passati alla terza fase della ricerca che qui verrà presentata.

Per verificare se le differenze di riconoscimento da parte dei rispondenti per ciascuna delle lingue e delle emozioni proposte nel test percettivo (*cfr.* § 10.5.2) siano da ricondurre alle diverse caratterizzazioni acustiche delle singole produzioni emotive presentate, è stata operata una analisi acustica delle produzioni utilizzate nel precedente *step*, al fine di estrarre una serie di indici acustici nel dominio della frequenza, dell'intensità e del tempo necessari a determinare i correlati acustici delle produzioni, e quindi delle emozioni, per ciascuna lingua considerata.

Lo scopo principale di questa fase della ricerca sarà, pertanto, orientato a:

1. caratterizzare e definire i profili acustici delle emozioni codificate in ciascuno degli stimoli in ciascuna delle lingue esaminate;
2. valutare se negli stimoli presentati agli ascoltatori e analizzati in questa fase vi siano differenze a livello acustico riconducibili e in grado di supportare le diverse percentuali di riconoscimento rilevate nella precedente indagine;
3. verificare se i dati che emergeranno dalla presente indagine trovano supporto nelle predizioni formulate da Scherer (1984, 1986, 2003; Banse, Scherer, 1996).

Dato l'elevato numero di emozioni considerate (complessivamente sette compresa la produzione neutra), complicata ulteriormente dalla presenza di più lingue, tutte le analisi verranno effettuate prendendo in esame solo gli *encoder* appartenenti alla tipologia degli *attori*, lasciando a ricerche future eventuali approfondimenti con i soggetti *naïf* che in questa sede sono stati utilizzati in via del tutto sperimentale: analizzando le matrici di confusione in Tabella 31 (*cfr.* § 10.5.2) è soprattutto per questi ultimi che si riscontrano maggiori elementi di confusione, dovuti molto probabilmente alla loro inadeguatezza nel produrre emozioni posate. La scelta di restringere le analisi ai soli *attori* consentirà non solo di circoscrivere il più possibile la discussione sulla tematica affrontata, ma si rende per lo più necessaria al fine di un raffronto con i dati e i risultati rinvenibili in letteratura, maggiormente disponibili per la tipologia dei soggetti *attori* che hanno avuto un impiego decisamente più ampio nelle ricerche sul parlato emotivo.

11.1 Estrazione dei dati

Come ampiamente trattato nel capitolo dedicato ai rapporti tra voce ed emozioni (*cfr.* capitolo 4) per l'individuazione dei correlati e delle proprietà acustiche nel parlato emotivo si è soliti eseguire misurazioni nel dominio della frequenza, dell'intensità e del tempo (*cfr.* tabella tratta da Scherer, Johnstone, Klassmeyer, 2003 e riportata in Appendice I).

Per gli indici nel dominio della frequenza (valori espressi in Hz) sono stati estrapolati i valori di F_0 massima ($F0_{max}$), minima ($F0_{min}$), media ($F0_{mean}$), deviazione standard ($F0_{stdev}$), mediana ($F0_{median}$), range ($F0_{range} = F0_{max} - F0_{min}$) e quartili 25 e 75 ($F0_{q25}$ e $F0_{q75}$).

Per gli indici nel dominio dell'intensità (valori espressi in dB) sono stati estrapolati i valori di intensità massima (Int_{max}), minima (Int_{min}), media (Int_{mean}), deviazione standard (Int_{stdev}), mediana (Int_{median}), range ($Int_{range} = Int_{max} - Int_{min}$) e quartili 25 e 75 (Int_{q25} e Int_{q75}).

Per gli indici nel dominio del tempo è stata misurata manualmente la durata in millisecondi (ms) dei due sintagmi che compongono gli stimoli (seg_1 e seg_2), la pausa (pause) tra i due sintagmi (ove presente) e il numero di sillabe presente nei due sintagmi necessario al calcolo dello *speech rate* (SR) secondo le indicazioni di Magno Caldognetto, Zmarich, Ferrero (1997) come numero di sillabe diviso per il tempo della catena fonica (syll./s). Nella schermata di seguito presentata in Figura 59 viene fornito un esempio di questa fase di lavoro effettuata in Praat da cui si evincono due *Tier*: sul primo *Tier* (segment) è stata effettuata la segmentazione e la misurazione della durata dei due sintagmi (seg_1 e seg_2); sul secondo *Tier* (pausesyll) è stato annotato il numero di sillabe presente all'interno e in corrispondenza dei due segmenti etichettati come seg_1 e seg_2 (di numero rispettivamente pari a 6 e 7 sillabe) e la durata della pausa (pause).

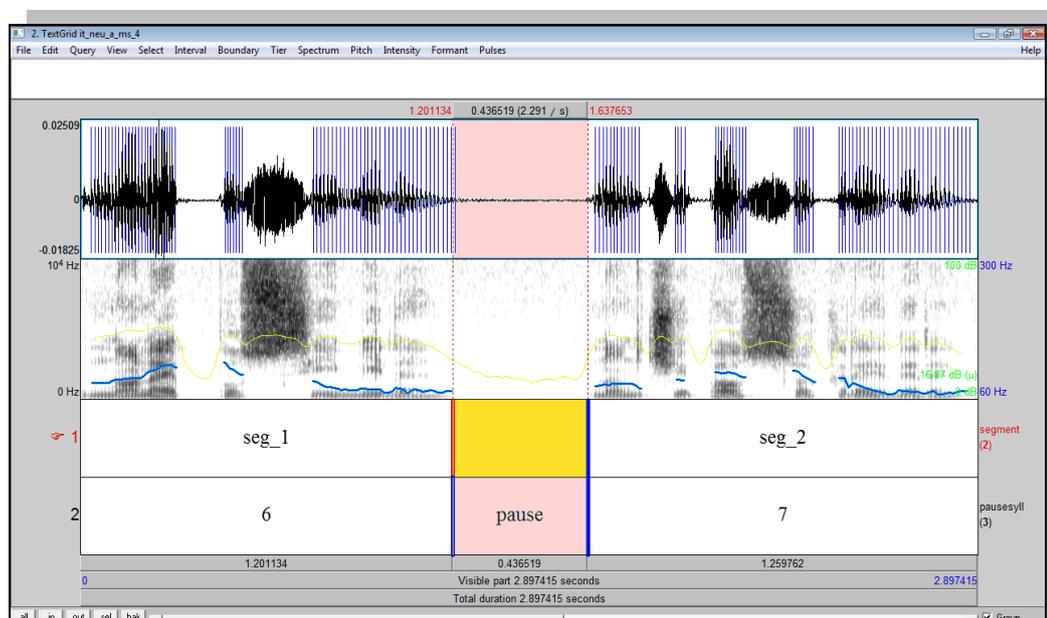


Figura 59: Esempio di schermata relativa alle misurazioni effettuate nel dominio del tempo

Poiché tutte le registrazioni sono state realizzate in camera silente (escludendo, quindi, la presenza di rumori o disturbi all'interno dei segnali registrati) e dato l'elevato livello di qualità delle stesse, si è optato per l'estrazione degli indici sopra elencati con l'ausilio di una procedura automatizzata mediante l'elaborazione e l'esecuzione di una routine all'interno del software Praat.³²³

³²³ Per le impostazioni e i dettagli dello *script* utilizzato si veda quanto riportato in Appendice XI.

Una volta effettuata l'estrazione di tutti i parametri, per semplificare le analisi comparative, si è proceduto al calcolo dei valori medi ottenuti sulla base dei dati relativi alle due porzioni di segnale (seg_1 e seg_2)³²⁴.

Le misurazioni, con riferimento ai valori mediati, estrapolate per i quattro *encoder attori*, ed utilizzate nel prosieguo dell'indagine qui condotta, vengono riportate in Tabella 45 in Appendice XII.

11.2 Analisi preliminare dei correlati acustici delle emozioni vocali³²⁵

In precedenti studi (si veda ad es. Magno Caldognetto *et al.*, 1998) è risultato che i valori minimi, massimi e *range* degli indici relativi a frequenza, intensità e durata sono correlati con i rispettivi valori medi. Nella presente indagine, per ciascuna lingua, per ciascuno dei parametri estrapolati per le 7 produzioni (6 emotive più la neutra) nel dominio di analisi relativo a frequenza e intensità, si è quindi proceduto al calcolo dell'indice di correlazione di *Pearson* mediante un'apposita funzione di SPSS per verificare se detti parametri fossero tra loro correlati con riferimento ai rispettivi valori medi³²⁶.

Dalle tabelle di correlazione riportate in Appendice XIII (rispettivamente Tabella 46 per gli indici estrapolati nel dominio della frequenza e Tabella 47 per quelli relativi al dominio dell'intensità) il dato riportato in Magno Caldognetto *et al.* (1998) risulta confermato nella presente indagine solo per gli indici legati ad F_0 : a parte i valori di $F0_min$ che in tutte e quattro le lingue risultano nel complesso scarsamente correlati con gli altri indici, tutti gli altri risultano significativamente correlati tra loro e con i valori di media e *range*. Si registra anche una correlazione tra media e *range* di F_0 che per l'italiano è di 0,961, per il francese di 0,525, per l'inglese di 0,767 e per il tedesco di 0,855. Come si evince dalla Tabella 47, non si hanno, invece, analoghi riscontri per gli indici nel dominio dell'intensità per i quali si registrano anche correlazioni negative.

Sebbene sia stata operata l'estrazione di tutti gli indici sopra menzionati, supportati dai risultati derivanti dal calcolo degli indici di correlazione tra i diversi parametri con i

³²⁴ In modo analogo si è proceduto per il calcolo dello *speech rate* medio (SR_mean), calcolato sui due segmenti (seg_1 e seg_2) aventi, rispettivamente, 6 e 7 sillabe l'italiano (vedi esempio in Figura 59), 3 e 5 il francese, 3 e 6 l'inglese e 3 e 5 il tedesco. Sebbene Kunzel (1997) faccia notare come l'utilizzo di catene foniche con un numero di sillabe inferiore a 6 contribuisca a far aumentare la variabilità intra-parlatore, nella presente indagine non ci si è posti questo limite e si è deciso di procedere comunque.

³²⁵ Si tiene a precisare che le analisi e le discussioni che seguiranno, data l'esiguità del campione analizzato (un soggetto *attore* per lingua), hanno carattere puramente descrittivo ed esplorativo volto alla ricerca di elementi di contatto o di discontinuità, a livello acustico, tra le emozioni prodotte in ciascuna lingua.

³²⁶ I valori medi a cui si fa riferimento sono i valori mediati dei due sintagmi seg_1 e seg_2 che sono stati inizialmente misurati separatamente (*cfr.* § precedente).

rispettivi valori di media e *range*, in questa sede è stata effettuata una valutazione dei soli indici macroprosodici riguardanti la media di F_0 ($F0_mean$), la variabilità o *range* di F_0 ($F0_range$), l'intensità media (Int_mean), il *range* dell'intensità (Int_range), la durata media e lo *speech rate* medio (SR_mean) per ciascuna emozione e per ciascun soggetto.

Sulla base di detti indici si è quindi proseguito nelle indagini quantificando e delineando, per ciascuna lingua, i profili acustici delle emozioni considerate.

11.2.1 Dominio della frequenza

Si osservi il grafico in Figura 60, sul quale vengono riportate le misurazioni effettuate per ciascuno stimolo di ciascuna lingua per la frequenza fondamentale F_0 , con particolare riferimento alla sua media e al suo *range* (dato da $F0_max - F0_min$).

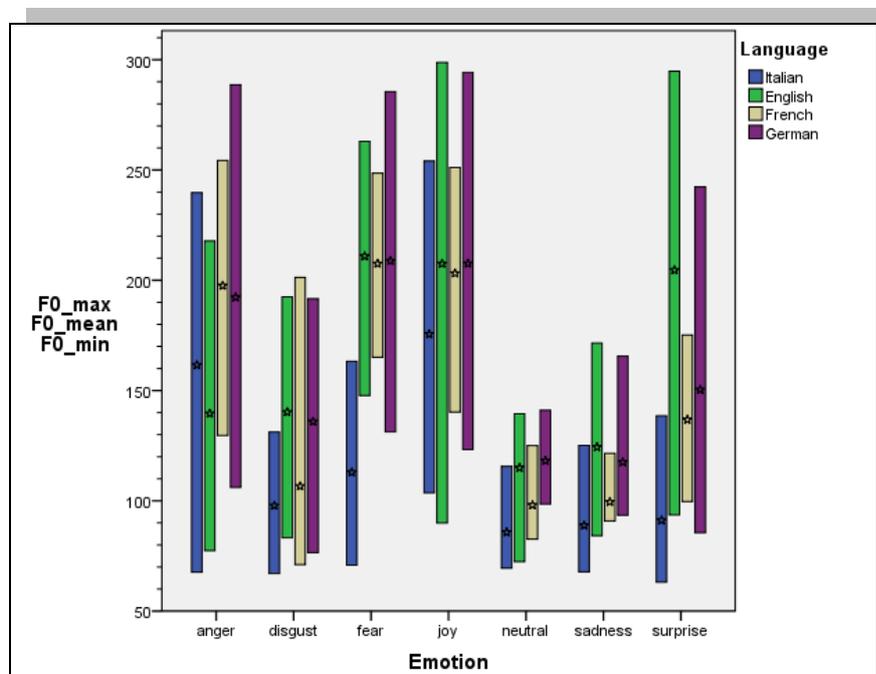


Figura 60: Distribuzione (in Hz) dei valori medi di $F0_max$, $F0_mean$ ed $F0_min$ per le sei emozioni più la produzione neutra prodotte da ciascun attore nelle propria lingua madre

Tralasciando per il momento l'analisi di F_0 media su cui si ritornerà più approfonditamente più avanti, il primo dato ad emergere è come sia proprio la produzione neutra, in tutte e quattro le lingue, ad essere la meno soggetta a fenomeni di variabilità di F_0 rispetto alle altre produzioni emotive. Il minore grado di variabilità ($F0_range$) tra le quattro lingue viene in tal caso rilevato per il tedesco (42,55 Hz) e per il francese (42,44 Hz), seguiti dall'italiano (con 46,19 Hz) e dall'inglese (con 67,05 Hz).

Per le emozioni di rabbia e gioia si registra in tutte e quattro le lingue la presenza di un *range* di F_0 più elevato rispetto, ad esempio, all'espressione della tristezza per la quale si

rilevano valori molto vicini a quelli della produzione neutra. Rispetto a quest'ultima per l'emozione di tristezza si ha un *range* leggermente inferiore per il francese (30,64 Hz), di poco superiore per l'italiano (57,29 Hz), maggiore di un terzo per l'inglese (87,29 Hz) e superiore di ca. 2/3 per il tedesco (72,13 Hz).

Per l'emozione di sorpresa in inglese e in tedesco si riscontra un *range* di F_0 (rispettivamente 201,18 Hz e 156,77 Hz), superiore di ca. tre volte il primo e di due volte il secondo rispetto a quello rilevato per l'italiano e il francese che hanno, invece, valori di *range* molto simili tra loro (rispettivamente 75,41 Hz e 75,52 Hz). Sempre per l'emozione di sorpresa in inglese si rileva, contrariamente alle altre lingue, anche un sensibile innalzamento di F_0 media). Riguardo l'emozione di paura italiana, rispetto all'inglese, al francese e al tedesco, si assiste ad uno spostamento verso il basso sia di F_0 media sia del *range* di F_0 . Anche per l'emozione di disgusto, se la si rapporta alla produzione neutra, si ha una F_0 media relativamente bassa pressoché in tutte e quattro le lingue.

Si evince, inoltre, come attraverso l'utilizzo del *range* di F_0 come indice macroprosodico (cfr. Figura 60), gli *encoder* di tutte e quattro le lingue abbiano marcato una netta contrapposizione tra emozioni ad alta attivazione (rabbia, gioia, sorpresa³²⁷) ed emozioni a bassa attivazione (tristezza e paura).

11.2.2 Dominio dell'intensità

Passando al dominio dell'intensità, esaminando i dati in Figura 61 si evince come in tutte e quattro le lingue solo l'emozione di tristezza viene prodotta con un'intensità media inferiore alla produzione neutra.

Il *range* dell'intensità si mantiene sufficientemente costante in tutte le lingue per tutte le emozioni, con l'inglese che in termini di variabilità dell'intensità si colloca indubbiamente al primo posto (soprattutto per la soglia di *Int_min* che è tra le più basse, a cui seguono in ordine crescente, l'italiano, il tedesco e il francese).

Si rileva, inoltre, una tendenza che vede la produzione di tutte le emozioni del tedesco (eccetto quella di rabbia) con un'intensità media tra le più alte rispetto alle altre lingue (a parte l'emozione di gioia del francese). Azzardando una possibile interpretazione, questa tendenza sembrerebbe spiegare l'elevato grado di confusione, da parte dei soggetti ascoltatori,

³²⁷ Ad eccezione dell'emozione di sorpresa dell'italiano e del francese, a cui si è già accennato, che non rispecchiano questa assunzione.

delle emozioni di paura, gioia, sorpresa e disgusto prodotte dall'attore tedesco con l'emozione di rabbia (*cfr.* Tabella 35 più sopra riportata).

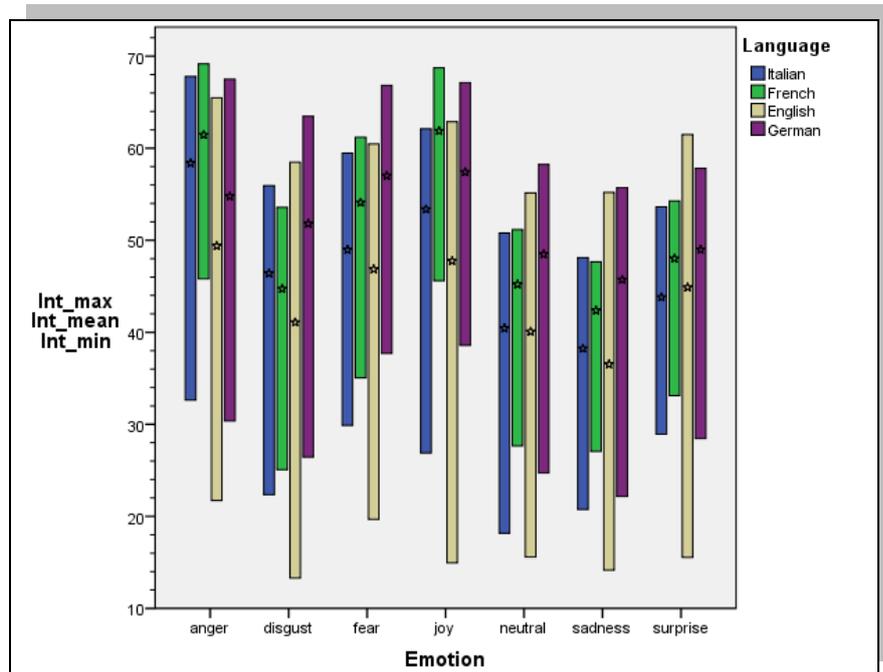


Figura 61: Distribuzione (in dB) dei valori medi di Int_max, Int_mean ed Int_min per le sei emozioni più la produzione neutra prodotte da ciascun attore nella propria lingua madre

11.2.3 Dominio del tempo

I dati qui presentati in riferimento alla durata assoluta delle singole produzioni nelle quattro lingue hanno, da un lato, uno scopo puramente indicativo, e dall'altro, grazie all'informazione ad essa connessa, è possibile mettere in evidenza un fattore di non poca importanza riconducibile alla diversa struttura sintattica e morfologica delle quattro lingue qui presentate.

Con riferimento al dominio del tempo il grafico in Figura 62 evidenzia chiaramente come gli stimoli prodotti in lingua tedesca abbiano in assoluto una durata media maggiore rispetto agli stimoli prodotti nelle altre lingue, seguiti solo da quelli prodotti in lingua italiana.

Esaminando il grafico in Figura 62 si può, tuttavia, dire che l'emozione di disgusto è stata prodotta dagli attori di tutte e quattro le lingue con un incremento di durata rispetto a tutte le altre emozioni, mentre per l'emozione di paura, se la si raffronta alle altre produzioni all'interno di ogni lingua, si rileva una riduzione dell'informazione connessa alla durata soprattutto nel francese, a cui seguono il tedesco e l'italiano (per l'inglese tale sistematicità non trova riscontro).

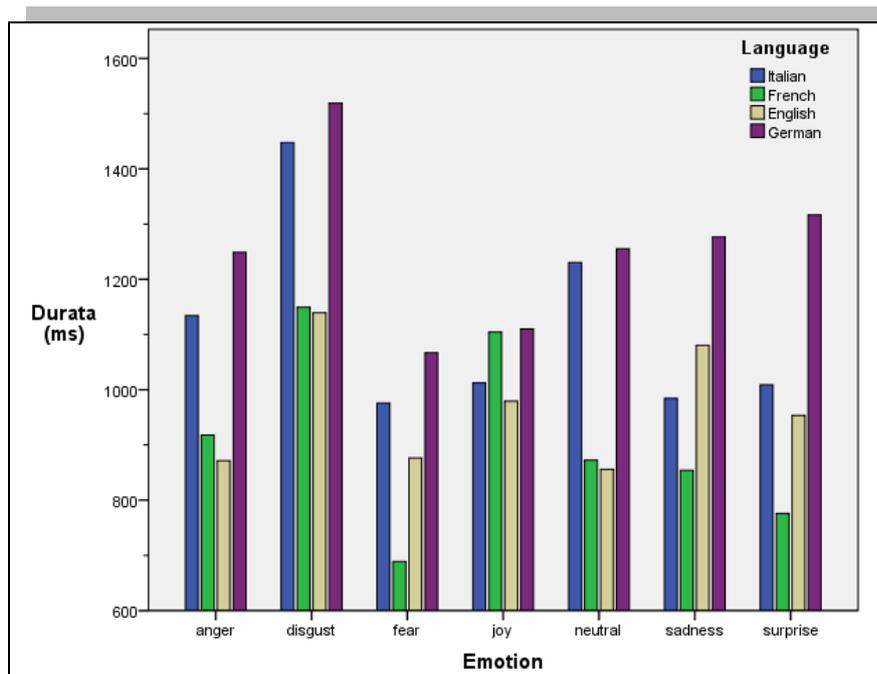


Figura 62: Durata media (in ms) degli enunciati prodotti dagli attori nella propria lingua madre per ciascuna emozione

Passando allo *speech rate* medio (SR_mean) calcolato per le produzioni qui prese in esame, gli stimoli prodotti dall'attore italiano si rivelano essere quelli col più alto *speech rate* (globalmente pari a 5,97 syll./s): i valori medi oscillano tra le 4,5 sillabe al secondo del disgusto e le 6,67 sillabe al secondo dell'emozione di paura (cfr. Figura 63). Anche nelle altre lingue, escluso l'inglese, l'emozione di paura è caratterizzata da un elevato *speech rate* con valori medi di 5,99 syll./s per il francese e di 5,16 syll./s per il tedesco.

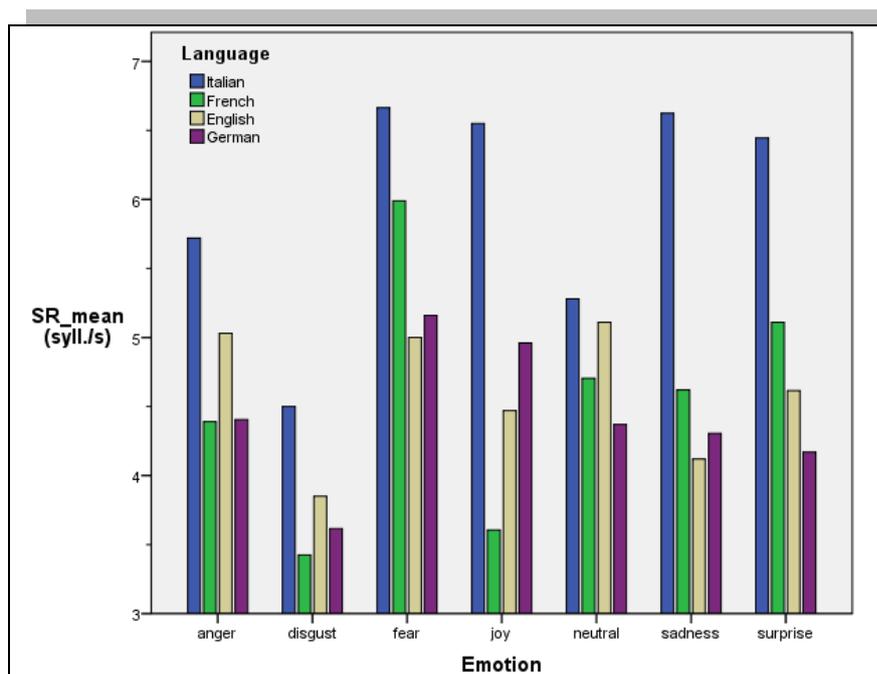


Figura 63: Andamento dello *speech rate* medio (syll./s) per ciascuna emozione prodotta dagli attori nella propria lingua madre

Per quel che concerne la pausa inserita tra i due sintagmi seg_1 e seg_2 si assiste ad un uso della stessa assai diversificato per i quattro soggetti analizzati (cfr. Figura 64).

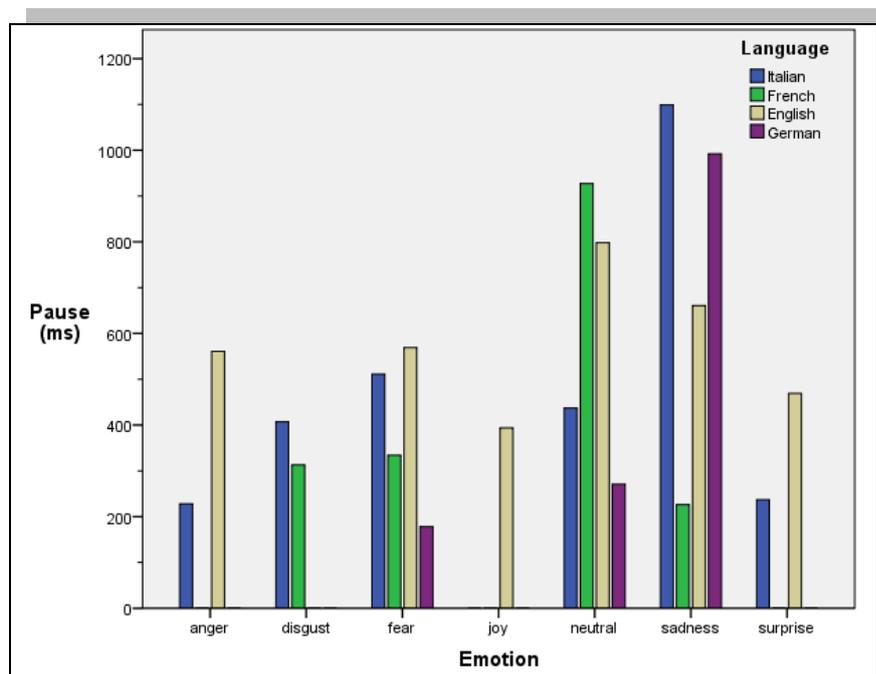


Figura 64: Durata della pausa (pause) inserita tra i due sintagmi (seg_1 e seg_2) da ciascun attore nella produzione degli enunciati nella propria lingua madre per ciascuna emozione

Tali differenze sono riconducibili, con buona probabilità, a diverse strategie comunicative e di pianificazione della produzione legate ai singoli soggetti. Da questo punto di vista una generalizzazione del fenomeno sarebbe del tutto fuorviante, non essendo lo stesso supportato da una casistica più ampia.

Tuttavia, analizzando i dati riportati in Figura 64, si evince come l'uso della pausa rappresenti una costante (sebbene con differenze legate agli aspetti appena richiamati) soprattutto nella produzione delle emozioni di paura e di tristezza, oltre che per la frase neutra. Nel complesso è l'italiano a fare maggiore ricorso all'uso della pausa tra i due sintagmi che raggiunge la sua durata massima (1099 ms) nella produzione dell'emozione di tristezza, mentre l'uso della stessa risulta azzerato nella produzione della gioia oltre che nell'italiano anche nel francese e nel tedesco.

11.3 Normalizzazione dei dati

Per poter meglio apprezzare le differenze tra i diversi parametri acustici estrapolati e le diverse configurazioni per ciascuna emozione, oltre che per ridurre la variabilità interparlatore, è stata eseguita una operazione di normalizzazione dei dati. Le procedure per la normalizzazione dei dati sono varie e diverse tra loro: in Anolli *et al.* (2008a, b), ad esempio,

dopo aver calcolato la differenza tra i valori della produzione emotiva e quelli della produzione neutra, la normalizzazione è stata ottenuta effettuando il rapporto tra detta differenza e i valori della produzione neutra; altri ancora (vedi ad esempio Forsell, 2007) hanno fatto ricorso a trasformazioni logaritmiche³²⁸.

Ai fini della presente indagine la normalizzazione dei dati è stata effettuata sottraendo ai valori originali (O) delle emozioni (E) un valore di riferimento (B, *baseline*), individuato nella produzione neutra (N) di ciascuno degli *encoder*, come nell'esempio che segue per il parametro F_0 :

$$\text{Norm } F_{0(E)} = OF_{0(E)} - BF_{0(N)}$$

In tal modo è possibile analizzare le diverse configurazioni di *patterns* acustici delle emozioni in termini di variazione in positivo o in negativo rispetto alla produzione neutra assunta come *baseline*.

11.4 Il profilo acustico delle emozioni nelle quattro lingue

Dopo un'analisi preliminare nel § 11.2 che ha consentito di trarre alcune considerazioni sui dati assoluti (valori medi non normalizzati) presentati per ciascuna lingua e per ciascuna emozione, grazie alla procedura di normalizzazione descritta nel precedente paragrafo, verranno qui definiti e messi a confronto i profili acustici delle sei emozioni nelle quattro lingue. L'analisi e la discussione dei dati verranno questa volta operate su valori relativi in quanto viene assunto come punto di riferimento la produzione neutra (o *baseline*, rappresentata dal punto zero nei grafici presentati), con il vantaggio di ridurre la variabilità interparlatore.

11.4.1 Le emozioni vocali in termini di frequenza

Come per l'analisi preliminare, in questa sede verranno analizzati i profili acustici delle emozioni nelle quattro lingue con riferimento ai valori, questa volta normalizzati, relativi a media e *range* di F_0 .

Si inizierà con l'analisi dei valori relativi alla F_0 media normalizzata, i cui dati vengono rappresentati graficamente in Figura 65. Osservando il grafico emerge come tutte e quattro le lingue ricorrano in modo analogo a variazioni di F_0 media (seppur con qualche

³²⁸ La normalizzazione è avvenuta attraverso la funzione logaritmica $\log(x/y) = \log(x) - \log(y)$.

differenza a cui si accennerà a breve) per produrre differenze nella produzione delle diverse espressioni emotive.

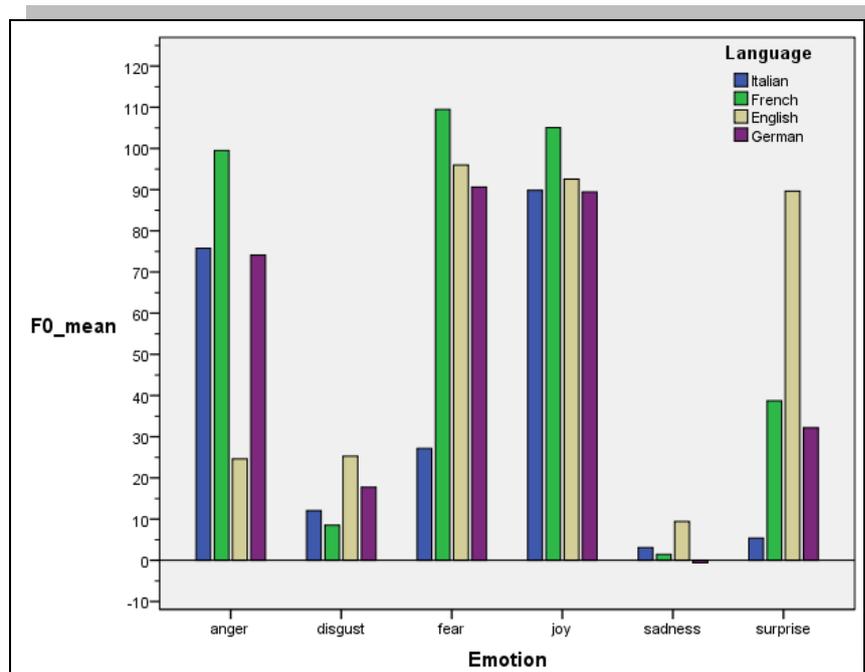


Figura 65: Variazione dei valori di F_0 media (F_0 _mean) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (*baseline*)

Nello specifico si rileva come le quattro lingue si comportino in modo pressoché simile nella produzione dell'emozione di gioia che viene realizzata con un sensibile incremento di F_0 media rispetto alla produzione neutra, dato che conferma l'alto grado di attivazione che contraddistingue questa emozione. Si rileva, inoltre, come i quattro attori si discostino di poco dai valori medi di F_0 nella produzione dell'emozione di tristezza che solo nel tedesco viene prodotta con un lievissimo decremento della F_0 media.

Un discorso a parte meritano le altre produzioni emotive, per le quali si registrano comportamenti diversi per ciascuna lingua da parte dei quattro attori. Ad esempio, con riferimento alla produzione neutra, l'emozione di rabbia viene prodotta dall'attore inglese con un evidente minore ricorso all'aumento di F_0 media (ca. 25 Hz) rispetto alle altre tre lingue per le quali si registrano, in rapporto all'inglese, valori triplicati per l'italiano e il tedesco e quadruplicati per il francese.

Sempre nell'inglese l'emozione di sorpresa viene prodotta con valori di F_0 media molto prossimi all'emozione di gioia, mentre per il francese e per il tedesco l'incremento della F_0 media rispetto alla produzione neutra è, rispettivamente, di ca. 39 Hz e di 32 Hz. Ciò non accade invece per l'italiano i cui valori sono prossimi a quelli della produzione neutra (con una differenza di ca. 5 Hz). Viene a questo proposito in mente quanto più sopra detto riguardo l'uso delle etichette emozionali nelle diverse lingue: dove può avere una sola valenza in una

lingua (senza distinzione tra stato positivo o negativo), può avere tale distinzione in un'altra lingua. Una possibile spiegazione potrebbe essere in questo caso data da quanto affermato da Härtel & Härtel (2005: 685-686):

“the word 'surprise' in English does not necessarily have a positive or negative connotation, whereas the German word 'Überraschung' probably has more of a positive than a negative connotation.”

È possibile, in questo caso, che lo scenario presentato per l'emozione di sorpresa sia stato interpretato dai diversi attori in modo diverso a seconda della lingua e la cultura di appartenenza: altamente positiva, quindi, quella prodotta dall'attore inglese, caratterizzata da un'elevata attivazione che spiegherebbe i valori di F_0 media molto vicini a quelli dell'emozione di gioia, mentre potrebbe essere stata interpretata con una valenza piuttosto negativa nel caso dell'italiano, con il tedesco e il francese che si collocano nella parte medio-bassa di questa scala di attivazione, o di positività/negatività.

Riguardo il *range* di variazione di F_0 si registrano, invece, maggiori differenze per ciascuna lingua (cfr. Figura 66).

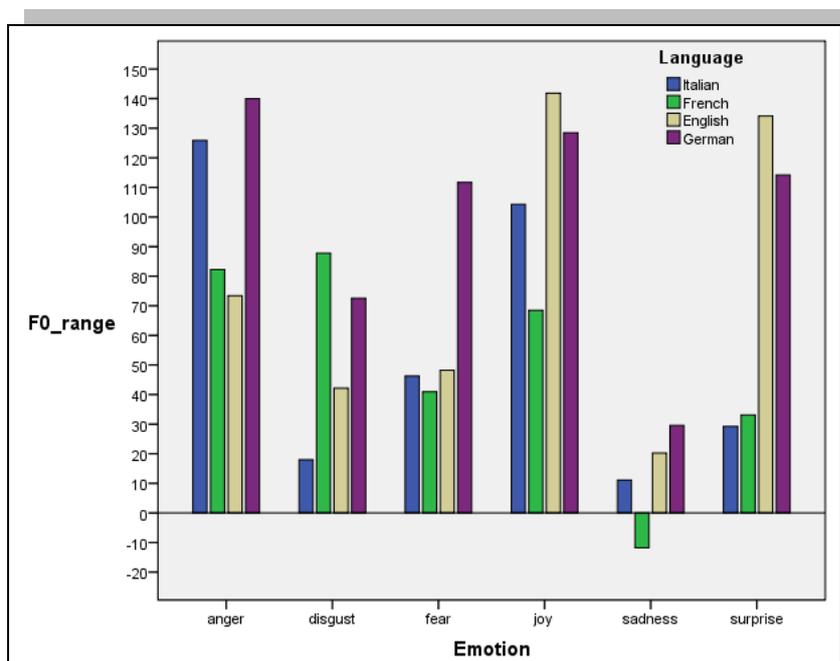


Figura 66: Variazione dei valori medi del range di F_0 (F_0_range) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (*baseline*)

Il *range* di F_0 risulta avere il valore assoluto più alto per l'emozione di rabbia prodotta dall'attore tedesco e italiano (anche rispetto alle altre emozioni nella medesima lingua).

L'emozione di disgusto viene prodotta con un *range* di variazione molto basso nell'italiano, mentre viene prodotta con un *range* maggiore nel francese (il maggiore in assoluto rispetto alle altre produzioni della stessa lingua).

L'emozione di paura ha un andamento comune, in termini di *range* di F_0 , per l'italiano, il francese e l'inglese, mentre viene realizzata con un *range* di poco più del doppio rispetto a queste ultime nell'attore tedesco.

Nell'emozione di gioia è l'inglese ad avere il *range* di F_0 più esteso seguito dal tedesco, dato che sembra riflettersi per le stesse, con una leggera riduzione, anche nella produzione dell'emozione di sorpresa. Il *range* di F_0 per l'emozione di paura è invece caratterizzato in tutte e quattro le lingue da valori molto bassi che si mantengono molto vicini a quelli della produzione neutra: solo nel francese si registra una evidente riduzione del *range* di F_0 rispetto alla produzione neutra.

Anche in questo caso a differenze nel *range* di F_0 corrisponde una distinzione tra emozioni ad alta attivazione (rabbia e gioia), caratterizzate da un *range* di F_0 maggiore, ed emozioni a bassa attivazione (tristezza), contraddistinte da un *range* di F_0 minore.

11.4.2 Le emozioni vocali in termini di intensità

Per l'intensità media (*cf.* Figura 67) si rileva una tendenza comune in tutte e quattro le lingue solo nella produzione delle emozioni di paura e di tristezza prodotte rispettivamente con un innalzamento e un abbassamento della stessa rispetto alla produzione neutra.

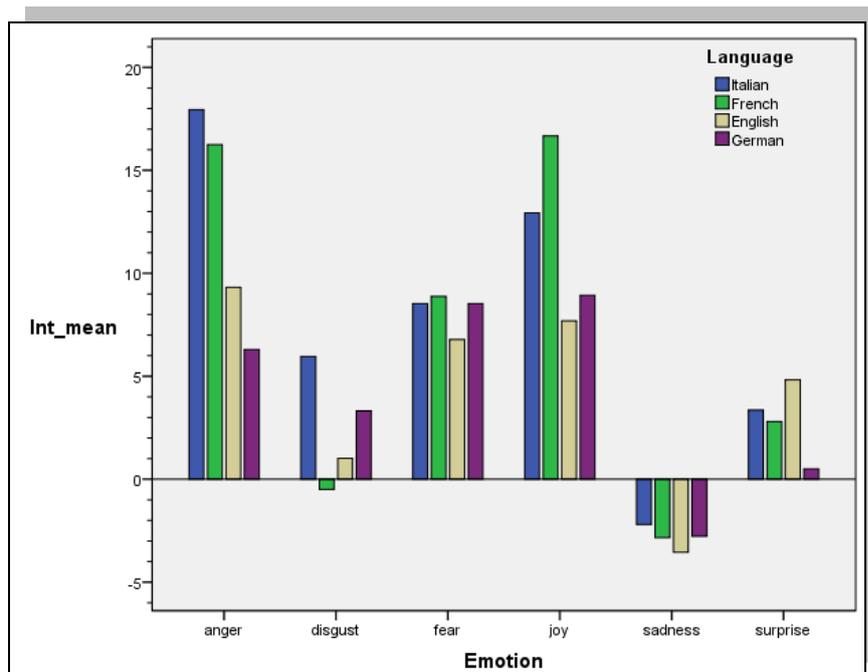


Figura 67: Variazione dei valori dell'intensità media (Int_mean) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (*baseline*)

Italiano e francese si avvicinano molto nell'espressione dell'emozione di rabbia, mentre per l'inglese e il tedesco si registrano valori che sono, rispettivamente, 1/2 e 1/3 dell'intensità media rilevata per l'italiano.

Praticamente nulla è invece la differenza per l'emozione di disgusto prodotta dall'attore francese e inglese, mentre si riscontra un aumento di ca. 6 dB per l'italiano e 3 dB per il tedesco.

Diversa è la situazione per l'emozione di gioia: la differenza rispetto alla produzione neutra è di ca. 13 dB nell'italiano, 17 dB nel francese, 9 dB nel tedesco e 8 dB nell'inglese.

Infine per l'emozione di sorpresa la differenza in rapporto alla produzione neutra è quasi pari a zero per il tedesco, mentre per le altre tre lingue si va dai ca. 3 dB dell'italiano e del francese ai quasi 5 dB dell'inglese.

Con riferimento al *range* dell'intensità, più sopra era stato rilevato come lo stesso si mantenesse sufficientemente costante in tutte le lingue per tutte le emozioni, con l'inglese che si collocava al primo posto per la maggiore estensione di questo indice. Analizzando i dati rappresentati in Figura 68 si rileva come vi sia, più che negli altri indici sin qui analizzati, una certa eterogeneità nell'uso del *range* dell'intensità nelle produzioni emotive che nelle quattro lingue assume valori sia positivi che negativi rispetto alla produzione neutra.³²⁹

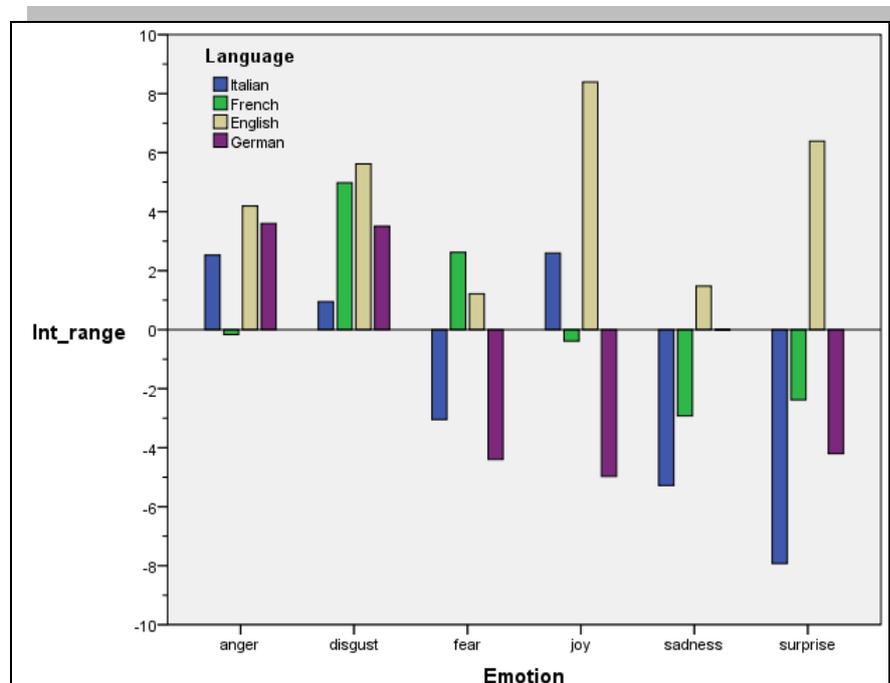


Figura 68: Variazione dei valori medi del *range* dell'intensità (Int_range) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (*baseline*)

³²⁹ Questo scarsa eterogeneità era già emersa dal calcolo degli indici di correlazione riportati in Tabella 47 e non si esclude che tale fenomeno possa essere attribuito alla diversa distanza dal microfono assunta dai soggetti in fase di produzione e raccolta delle frasi (sebbene vada sottolineato come quest'aspetto sia stato fortemente limitato in origine durante le fasi di raccolta del materiale proprio per evitare situazioni simili).

11.4.3 Le emozioni vocali in termini di durata

Dopo aver ridotto, grazie alla normalizzazione, la differenza di durata insita nella struttura e nella tipologia delle quattro lingue esaminate, analizzando il grafico in Figura 69 si evince come solo nel caso dell'emozione di disgusto i quattro attori ricorrano in modo omogeneo ad un aumento della durata.

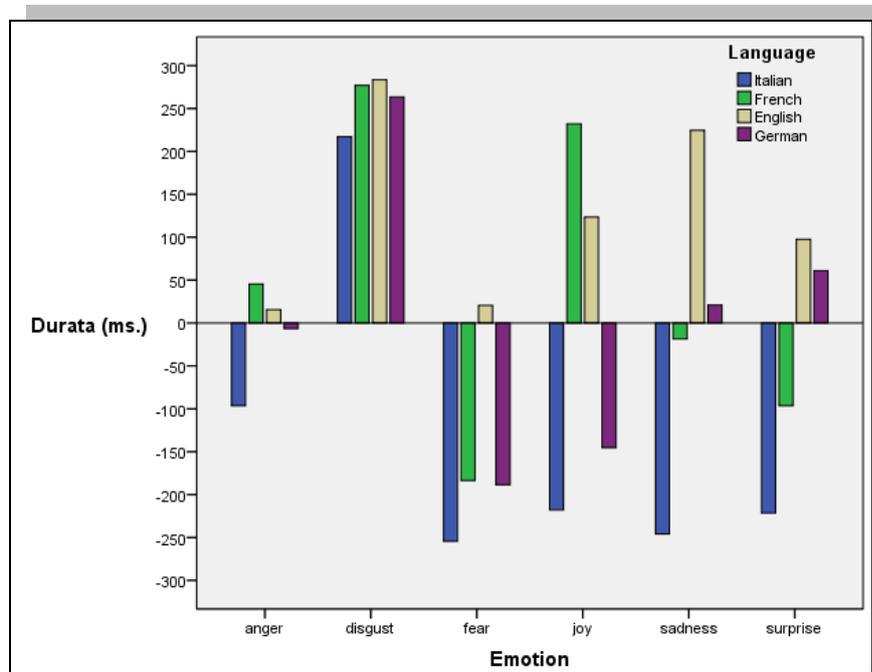


Figura 69: Variazione dei valori medi di durata per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (*baseline*)

L'attore italiano tende in generale ad una compressione del tempo nella produzione di tutte le frasi emotive.³³⁰ Fatto analogo si verifica per l'attore tedesco limitatamente alle produzioni di paura e gioia, mentre per le emozioni di rabbia, tristezza e sorpresa la variazione di durata in positivo può ritenersi trascurabile.

È invece l'attore inglese a fare maggiore utilizzo dell'informazione connessa alla durata, per il quale si registrano variazioni in positivo rispetto alla frase neutra soprattutto nella produzione delle emozioni di gioia, tristezza e sorpresa, mentre la variazione (sempre in positivo) risulta sufficientemente trascurabile per le emozioni di rabbia e paura.

Rispetto alla frase neutra, per l'attore francese si registra, invece, un evidente ricorso all'aumento della durata soprattutto per l'emozione di gioia, mentre si verifica il contrario per le emozioni di paura e sorpresa; per le emozioni di rabbia e tristezza la variazione, positiva nella prima, negativa nella seconda, sono anche in questo caso trascurabili.

³³⁰ È fin troppo evidente che nel produrre la frase neutra vi sia stata, da parte di tutti i soggetti, un'inconscia iperarticolazione che nel caso dell'italiano risulta maggiormente accentuata.

Per quanto riguarda invece lo *speech rate* medio rispetto alla produzione relativa alla frase neutra si registra, come è normale che sia, una rapporto inversamente proporzionale a quello della durata: ad un aumento della durata corrisponde una riduzione dello *speech rate* e, viceversa, alla diminuzione della durata corrisponde un aumento dello *speech rate*. Il grafico in Figura 70 viene riportato solo a titolo esaustivo.

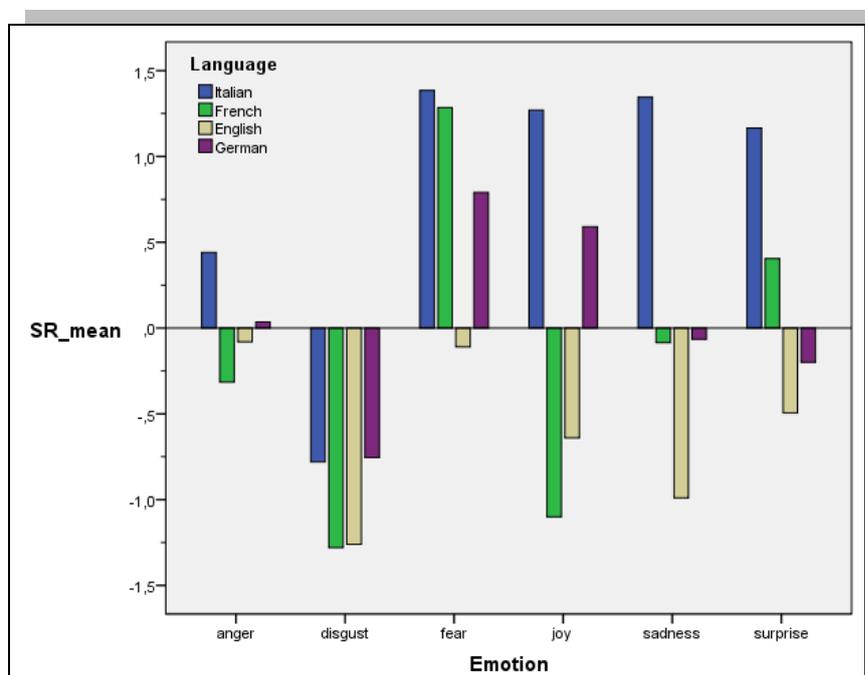


Figura 70: Variazione dei valori medi di *speech rate* (SR_mean) per ciascuna emozione e lingua normalizzati rispetto alla produzione neutra (*baseline*)

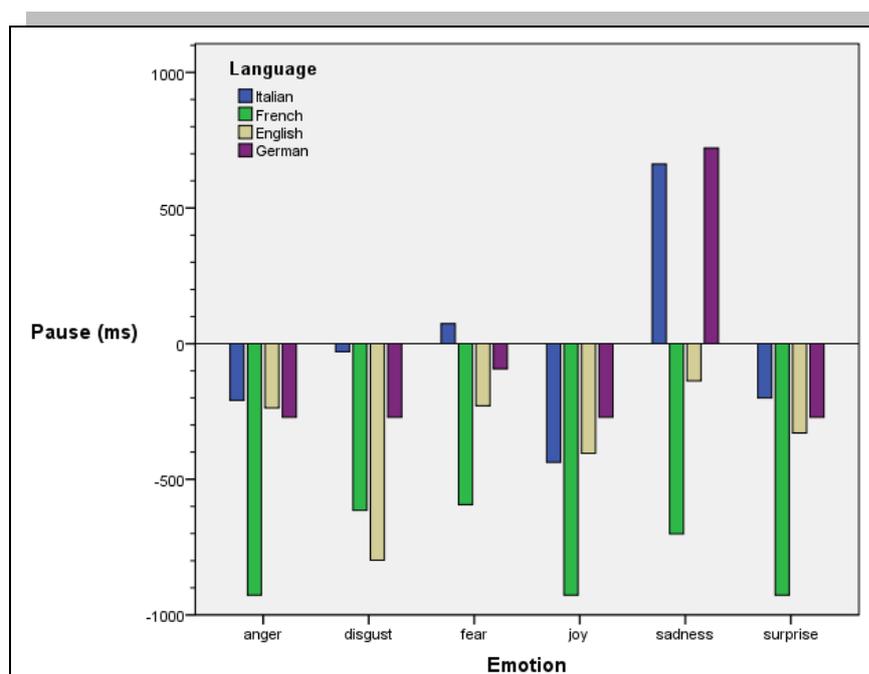


Figura 71: Utilizzo delle pause per ciascuna emozione e lingua normalizzate rispetto alla produzione neutra (*baseline*)

Per concludere, infine, era già risultato evidente come nel complesso vi fosse una sostanziale riduzione delle pause in tutte le produzioni emotive rispetto alla produzione della frase neutra, riduzione che in alcuni casi portava addirittura ad un totale annullamento della stessa. Solo l'italiano e il tedesco ricorrono ad un sostanziale allungamento della pausa nella produzione dell'emozione di tristezza (*cfr.* Figura 71 riportata sopra).

12 Conclusione

Bibliografia

- Abelin, Å., Allwood, J. (2000), Cross linguistic interpretation of emotional prosody, in: *SpeechEmotion-2000*, pp. 110-113.
- Abelin, Å., Allwood, J. (2002), Cross linguistic interpretation of emotional prosody, *Papers in theoretical linguistics*, Gothenburg, pp. 1-18.
- Albano Leoni, F., Maturi, P. (2003), *Manuale di Fonetica*, Roma, Carocci Editore.
- Albas, D. C., McCluskey, K. W., Albas, C. A. (1976), Perception of the emotional content of speech - A comparison of two Canadian groups, *Journal of Cross-Cultural Psychology*, 7: 481-489.
- Amrus, D. C. (2000), Collecting and recording of an emotional speech database, *Technical Report*, Faculty of Electrical Engineering, Institute of Electronics, Univ. of Maribor.
- Amir, N., Ziv, S., Cohen, R. (2003), Characteristics of authentic anger in hebrew speech, in *EUROSPEECH-2003*, pp. 713-716.
- Anolli, L. (2006), *Fondamenti di psicologia della comunicazione*, Bologna, Il Mulino.
- Anolli, L., Ciceri, R. (1992), *La voce delle emozioni. Verso una semiosi della comunicazione vocale non-verbale delle emozioni*, Milano: Angeli.
- Anolli, L., Wang, L., Mantovani, F. & De Toni, A. (2008a), La voce delle emozioni in giovani adulti cinesi e italiani, in: *Atti del 1° Convegno GSCP. Comunicazione parlata e manifestazione delle emozioni*, a cura di E. Magno Caldognetto, F. Cavicchio e P. Cusi, Napoli: Liguori Editore, pp. 2-44.
- Anolli, L., Wang, L., Mantovani, F. & De Toni, A. (2008b), The Voice of Emotion in Chinese and Italian Young Adults, in: *Journal of Cross-Cultural Psychology* 39: 565-598.
- Arnold, M. B. (1960), *Emotion and personality Psychological aspects*, (Vol. 1), New York: Columbia University Press.
- Averill, J. R. (1980), A constructivist view of emotion, in R. Plutchik e H. Kellerman (Eds.), *Theories of emotion*, New York: Academic Press.

- Averill, J. R. (2004), Everyday Emotions: Let Me Count the Ways, *Social Science Information* 43(4): 571-80.
- Ax, A. F. (1953), The physiological differentiation between fear and anger in humans, *Psychosomatic Medicine*, 15(5): 433-442.
- Balconi, M. (2004), *Neuropsicologia delle emozioni*, Roma, Carocci.
- Baños, R., Liaño, V., Botella, C., Alcañiz, M., Guerrero, B., Rey, B. (2006), Changing induced moods via virtual reality, in: *PERSUASIVE 2006, LNCS 3962*, pp. 7-15.
- Banse, R. & Scherer, K. R. (1996), Acoustic Profiles in Vocal Emotion Expression, *Journal of Personality and Social Psychology*, 170: 614-636.
- Barbaranelli, C. & D'Olimpo, F. (2007), *Analisi dei dati con SPSS. Vol. I: Le analisi di base*, Milano, LED.
- Batliner, A., Hacker, C., Steidl, S., Nöth, E., D'Arcy, S., Russel, M., & Wong, M. (2004). "You stupid tin box" - children interacting with the AIBO robot: a cross-linguistic emotional speech corpus, in: *Proceedings of the 4th International Conference of Language Resources and Evaluation (LREC '04)*, pp. 171-174.
- Battacchi, M. W., Renna, M., Suslow, Th. (1995), *Emozioni e linguaggio*, Roma, La Nuova Italia Scientifica.
- Beier, E., Zautra, A. (1972), Identification of vocal communication of emotions across cultures, in: *Journal of Consulting and Clinical Psychology*, 39(1): 166.
- Braun, A., Oba, R. (2007), Speaking Tempo in Emotional Speech – a Cross-Cultural Study Using Dubbed Speech, in: *Proceedings of the International workshop on Paralinguistic Speech - between models and data, ParaLing'07*, 3 August 2007, Saarbrücken, Germany, <http://www2.dfki.de/paraling07/papers/16.pdf>
- Breitenstein, C., Van Lancker, D., Daum, I. (2001), The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample, in: *Cognition & Emotion* 15(1): 57-79.
- Burkhardt, F., Audibert, N., Malatesta, L., Turk, O., Arslan, L., Auberge, V. (2006), Emotional prosody - does culture make a difference?, *Proceedings of Speech Prosody 2006, Dresden, Germany*, paper 207.
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., Weiss, B. (2005), A database of German emotional speech, in *INTERSPEECH-2005*, pp. 1517-1520.

- Campbell, N. (2000), Databases of emotional speech, in *SpeechEmotion-2000*, pp. 34-38.
- Canepari, L. (1985), *L'intonazione, linguistica e paralinguistica*, Napoli, Liguori editore.
- Chung, S. J. (1999), Vocal expression and perception of emotion in Korean, in: *Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, USA, pp. 969-972.
- Chung, S. J. (2000), *L'expression et la perception de l'émotion extraite de la parole spontanée: évidences du coréen et de l'anglais*, Unpublished doctoral dissertation, Université de la Sorbonne Nouvelle, Paris III, France, <<http://www.geocities.com/soojinchung/Finalthesis.pdf>>.
- Clark, D. M. (1983), On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures, in: *Advanced Behavior Research and Therapy*, 5: 27-49.
- Clore, G. L. & Ortony, A. (2000), Cognition in emotion: Always, sometimes, or never?, in: L. Nadel, R. Lane & G. L. Ahern (Eds), *The Cognitive neuroscience of emotion*, New York: Oxford University Press, pp. 24-61.
- Coan, J. A. & Allen, J. J. B. (2007), *Handbook of Emotion Elicitation and Assessment*, New York, NY: Oxford University Press.
- Cornelius, R. (1996), *The science of emotion*, New Jersey: Prentice Hall.
- Cosi P., Fusaro A., Tisato G. (2003), "LUCIA a New Italian Talking-Head Based on a Modified Cohen-Massaro's Labial Coarticulation Model", Proc. Eurospeech 2003, Geneva, Switzerland, pp. 127-132.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J. (2001), Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine*, 18: 32-80.
- Cronbach, L. J. (1951), Coefficient alpha and the internal structure of tests, *Psychometrika* 16(3), University of Illinois.
- D'Urso, V., Trentin, R. (2004), *Introduzione alla Psicologia delle emozioni*, Roma, Laterza.
- Damasio, A. R. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, Grosset/Putnam, New York.
- De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M. (2003), APML, a mark-up language for believable behavior generation, in: Prendiger, H., Ishizuka, M. (Eds.), *Life-like*

- characters. Tools, affective functions and applications*, Berlin: Springer-Verlag, pp. 65-86.
- De Saussure, F. (1972), *Cours de linguistique générale*, Paris, Edition Payot.
- De Sousa, R. (1987), *The Rationality of Emotion*, Cambridge MA: The MIT Press.
- Douglas-Cowie, E., Campbell, N., Cowie, R. & Roach, P. (2003), Emotional speech: Towards a new generation of databases, *Speech Communication*, 40: 33–60.
- Dromey, C., Silveira, J., Sandor, P. (2005), Recognition of affective prosody by speakers of English as a first or foreign language, in: *Speech Communication* 47(3): 351-359.
<http://dx.doi.org/10.1016/j.specom.2004.09.010>
- Eich, E., Ng, J. T. W., Macaulay, D., Percy, A. D., & Grebneva, I. (2007), Combining music with thought to change mood, in J. A. Coan & J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment*, New York: Oxford University Press, pp. 124-136.
- Ekman, P. (1992), An argument for basic emotions, *Cognition and Emotion*, 6, pp. 169-200.
- Elfenbein, H. A., & Ambady, N. (2002), On the universality and cultural specificity of emotion recognition: A meta-analysis, *Psychological Bulletin*, 128: 203-235.
- Erickson, D. & Maekawa, K. (2001), Perception of American English emotion by Japanese listeners, *Acoustical Society of Japan, Spring Meeting*, pp. 333-334.
- Fagyal, Z. (1995), *Aspects phonostylistiques de la parole médiatisée lue et spontanée: Age, prestige, situation, style et rythme de parole de l'écrivain M. Duras*, Thèse de Doctorat, Université de la Sorbonne Nouvelle, Paris.
- Fairbanks, G. & Hoaglin, L. W. (1941), An experimental study of the durational characteristics of the voice during the expression of emotion, *Speech Monograph*, 8: 85-91.
- Fairbanks, G. & Pronovost, W. (1939), An experimental study of the pitch characteristics of the voice during the expression of emotion, *Speech Monograph*, 6: 87-104.
- Fant, G. (1960), *Acoustic theory of speech production*, Mouton, The Hague.
- Fehr, B., Russell, A. J. (1984), Concept of emotion viewed from a prototype perspective, *Journal of Experimental Psychology: General*, 113(3): 464-486.

- Ferrero, F. E., Magno Caldognetto, E. (1986), Elementi di fonetica acustica, in *Trattato di Foniatria e Logopedia*, (L. Croatto, a cura di), Padova, La Garangola, pp. 155-196.
- Forsell, M. (2007), Acoustic correlates of perceived emotions in speech, Stockholm: *Numerisk analys och datalogi*, Kungliga Tekniska Högskolan.
- Gerrards-Hesse, A., Spies, K., & Hesse, E. W. (1994), Experimental inductions of emotional states and their effectiveness: A review, *British Journal of Psychology*, 85: 55-78.
- Gola, E. (2002), L'approccio alle emozioni in IA, in Bazzanella, C., Kobau, P. (a cura di), *Passioni, emozioni, affetti*, Milano: McGraw-Hill, pp. 177-196.
- Gonzalez, G. M. (1999), Bilingual computer-assisted psychological assessment: an innovative approach for screening depression in Chicanos/Latinos, *Technical Report 39*, Univ. Michigan.
- Goodman, R., "William James", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), URL = < <http://plato.stanford.edu/entries/james/>>
- Graham, C.R., Hamblin, A., Feldstein, S. (2001), Recognition of emotion in English voices by speakers of Japanese, Spanish and English, *IRAL*, 39: 19-37.
- Harré, R. (1986), *The social construction of emotions*, Oxford: Blackwell.
- Härtel, C. E. J. & Härtel, G. F. (2005), Cross-cultural differences in emotions: the why and how, *Social Science Information* 44(4): 683-693.
- Helfrich, H., Standke, R. & Scherer, K. R. (1984), Vocal indicators of psychoactive drug effects, in: *Speech communication* 3: 245-252.
- Hohmann, G. W. (1966), Some Effects of Spinal Cord Lesions on Experienced Emotional Feelings, in: *Psychophysiology*, 3: 143-56.
- Hollien, H. (2002), *Forensic voice identification*, Academic Press.
- Irwin, J. R. (1947), Galen on the Temperaments, *Journal of General Psychology*, 36: 45-64.
- Jakobson, R. (1963), *Essais de linguistique générale*, Paris, Minuit, (trad. it. *Saggi di linguistica generale*, Milano, Feltrinelli, 1966).
- James, W. (1884), "What Is An Emotion?", *Mind* 9: 188-205.
- Johnson, K. A. (2005), Speaker normalization in speech perception, in: D. B. Pisoni and R. E. Remez (Eds.), *The handbook of speech perception*, Blackwell Publishing, Malden, MA, pp. 363-389.

- Johnstone, T. & Scherer, K. R. (2000), Vocal Communication in Emotion, in: *Handbook of Emotion* (M. Lewis & J. Haviland, eds.), New York: Guilford, pp. 220-235.
- Johnstone, T., van Reekum, C. M., Hird, K., Kirsner, K., Scherer, K. R. (2005), Affective speech elicited with a computer game, *Emotion*, 5: 513-518.
- Juslin, P. N. & Laukka, P. (2003), Communication of emotions in vocal expression and music performance: Different channels, same code?, *Psychological Bulletin*, 129: 770-814.
- Kenealy, P. (1986), The Velten Mood Induction Procedure: A Methodological Review, *Motivation and Emotion*, 10(4): 315-335.
- Kienast, M., Paeschke, A. & Sendlmeier, W. F. (1999), Articulatory reduction in emotional speech, *Proceedings Eurospeech 99*, Budapest, Vol. 1, pp. 117-120.
- Klingholz, F. & Martin, F. (1985), Quantitative spectral evaluation of shimmer and jitter, *J. Speech Hear. Res.*, 28: 169-174.
- Kori, S., Magno Caldognetto, E. (2003), La caratterizzazione fonetica delle emozioni: primi dati da uno studio cross-linguistico italiano-giapponese, in: *Voce-Canto-Parlato. Studi in onore di Franco Ferrero* (P. Cosi, E. Magno Caldognetto & A. Zamboni, a cura di), pp.187-200. Padova: Unipress.
- Kramer, E. (1964), Elimination of verbal cues in judgments of emotion from voice, in: *Journal of Abnormal and Social Psychology*, 68(4): 390-396.
- Kunzel, H. J. (1997), Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, 4(1): 48-83.
- Ladd, D. R. & Terken, J. (1995), Modelling intra-and interspeaker pitch range, in: *Proceedings of ICPhS '95*, Stockholm, Vol. 2: 386-9.
- Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G. & Scherer, K. R. (1985), Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect, *Journal of the Acoustical Society of America*, 78: 435-444.
- Ladefoged, P. (1967), *Three areas of experimental phonetics*, Oxford University Press.
- Ladefoged, P. (1975), *A course in phonetics*. Orlando: Harcourt Brace, (5th Edn., Boston: Thomson Wadsworth, 2006).

- Larsen, R. J., & Sinnett, L. M. (1991), Meta-analysis of experimental manipulations: Some factors affecting the Velten Mood induction procedure, in: *Personality and Social Psychology Bulletin*, 17(3): 323–334.
- Laukka, P. (2004), Vocal expression of emotion: discrete-emotions and dimensional accounts, *Ph.D thesis*, Uppsala University.
- Laukka, P., Elenius, K., Fredriksson, M., Furumark, T., & Neiberg, D. (2008), Expression in spontaneous and experimentally induced affective speech: Acoustic correlates of anxiety, irritation and resignation, in: *Workshop on Corpora for Research on Emotion and Affect*, Marrakesh, Marocco.
- Laver, J. (1980), Phonatory settings, *The Phonetic Description of Voice Quality*, Cambridge: Cambridge University Press, pp. 93-135.
- Laver, J. (1994), *Principles of Phonetics*, Cambridge: Cambridge University Press.
- Lazarus, R. S. (1966), *Psychological Stress and the Coping Process*. New York: McGraw-Hill.
- Lazarus, R. S. (1982), Thoughts on the relations between emotion and cognition, *American psychologist*, 37(9): 1019-1024.
- Lazarus, R. S. (1984), On the Primacy of Cognition, *American Psychologist*, 39(2): 124-129.
- LeDoux, J. E. (1996), *The Emotional Brain*, New York: Touchstone.
- Leoni, A., Maturi, P. (2002), *Manuale di fonetica*, Roma, Carocci Editore.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. S., Studdert-Kennedy, M. (1967), Perception of the speech code, in *Psychological Review*, 74: 431-461.
- Liberman, A. M., Mattingly, I. G. (1985), The motor theory of speech perception revised, *Cognition* 21: 1-36.
- Lieberman, P. (1961), Perturbations in vocal pitch, *J. Acoust. Soc. Am.* 33: 597-603.
- Magno Caldognetto E., Zmarich C., Ferrero E. (1997), A comparative acoustic study of spontaneous and read Italian speech, *Proceedings of "5th European Conference on Speech Communication and Technology" Eurospeech '97* (Rhodes-Greece, 22-25 Sept. 1997), pp. 779-782.
- Magno Caldognetto, E. (2002), I correlati fonetici delle emozioni, in: *Passioni, emozioni, affetti* (C. Bazzanella & P. Kobau, a cura di), Milano: McGraw-Hill, pp. 197-213.

- Magno Caldognetto, E., Cavicchio, F., Cosi, P. (2008), La faccia e la voce delle emozioni, in: Poggi I. (a cura di), *La mente del cuore. Scienze cognitive ed emozioni*, Roma: Armando, pp. 178-212.
- Magno Caldognetto, E., Cavicchio, F., Cosi, P., Drioli, C., Tisato, G. (2005), Parametri per lo studio delle modificazioni articolatorie nel parlato emotivo, in Cosi P. (a cura di), *CD-Rom Proceedings of AISV 2004, 1st Conference of Associazione Italiana di Scienze della Voce, Padova, Italy, December 2-4, 2004*, Padova, EDK Editore s.r.l., pp. 449-470.
- Magno Caldognetto, E., Cosi, P., Cavicchio, F. (2004a), Modifications of Speech Articulatory Characteristics in the Emotive Speech, in *Proceedings of Tutorial and Research Workshop Affective Dialogue Systems 2004*, June 14 – 16, Kloster Irsee (Germany), pp. 106-112.
- Magno Caldognetto, E., Cosi, P., Cavicchio, F., Drioli, C., Tisato, G. (2003a), Modificazioni Articolatorie ed Acustiche Dovute alla Coproduzione di Emozioni e Parlato, in *Atti delle XIV giornate del GFS Costituzione, Gestione e Restauro di Corpora Vocali*, Viterbo, 4-6 Dicembre, pp. 234-341.
- Magno Caldognetto, E., Cosi, P., Drioli, C., Tisato, G., Cavicchio, F. (2003b), Coproduction of Speech and Emotions: Visual and Acoustic Modifications of Some Phonetic Labial Targets, in *Proceedings of AVSP 03*, September 4-7, St. Jorioz (France), pp. 209-214.
- Magno Caldognetto, E., Cosi, P., Drioli, C., Tisato, G., Cavicchio, F. (2004b), Modifications of phonetic labial targets in emotive speech: Effects of the coproduction of speech and emotions, *Speech Communication*, 44: 173-185.
- Magno Caldognetto, E., Kori, S. (1983), Intercultural judgment of emotions expressed through voice, in: *Quaderni del Centro di Studio per le Ricerche di Fonetica*, 2: 339-363.
- Magno Caldognetto, E., Zmarich, C., Ferrero, F. E. (1998), Indici acustici macroprosodici dello stato emotivo del parlante, in *Atti del XXVI Convegno Nazionale di Acustica*, pp. 263-268.
- Mandler, G. (1990), A constructivist theory of emotion, in N. S. Stein, B. L. Leventhal and T. Trabasso (Eds.), *Psychological and biological approaches to emotion*, Hillsdale, N.J.: Lawrence Erlbaum Associates.

- Massaro, D. W., Cole, R. (2000), From Speech is special to talking heads in language learning, in: *Proceedings of Integrating Speech Technology in the (Language) Learning and Assistive Interface (InSTIL 2000)*, University of Abertay Dundee, Scotland, pp. 153-161
- Matthei, E., Roeper, T. (1991), *Elementi di psicolinguistica*, Bologna, il Mulino.
- Maturi, P. (2006), *I suoni delle lingue, i suoni dell'italiano*, Bologna, Il Mulino.
- McCluskey, K.W., Albas, D. C. (1981), Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults, *International Journal of Psychology*, 16: 119-132.
- McCluskey, K.W., Albas, D. C., Niemi, R. R., Cuevas, C., Ferrer, C. A. (1975), Cross-cultural differences in the perception of emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children, *Developmental Psychology*, 11: 15-21.
- Mehrabian, A. (1971), *Silent messages*, Belmont, CA: Wadsworth.
- Mehrabian, A., Ferris, S. R. (1967), Inference of attitudes from nonverbal communication in two channels, *Journal of consulting psychology* 31(3): 248-252.
- Mehrabian, A., Russell, J. A. (1974), *An Approach to Environmental Psychology*. MIT Press, Cambridge, MA, USA; London, UK.
- Mehrabian, A., Wiener, M. (1967), Decoding of inconsistent communications, *Journal of personality and social psychology* 6(1): 109-114.
- Mesquita, B. & Leu, J. (2007), The Cultural Psychology of Emotion, in S. Kitayama and D. Cohen (Eds.), *The Handbook of Cultural Psychology*, New York: Guilford Press.
- Mesquita, B. (2007), Emotions are culturally situated, *Social Science Information*, 46(3): 410-415.
- Mioni, A. (1986), Fonetica articolatoria: descrizione e trascrizione degli atteggiamenti articolatori, in *Trattato di Foniatria e Logopedia*, (L. Croatto, a cura di), Padova, La Garangola, pp. 15-89.
- Nakamichi, A., Jogan, A., Usami, M. & Erickson, D. (2002), Perception by native and non-native listeners of vocal emotion in a bilingual movie, in: *Gifu City Women's College Research Bulletin*, 52: 87-91.

- Niedenthal, P. M., Krauth-Gruber, S., Ric, F. (2006), *The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches*, Principles of Social Psychology series, New York: Psychology Press.
- Oatley, K. (2007), *Breve storia delle emozioni*, ed. italiana a cura di L. Anolli, Universale Paperbacks Il Mulino, Bologna.
- Ortony, A., Clore, G. L., Collins, A. (1988), *The Cognitive Structure of Emotion*, UK: Cambridge University Press.
- Osgood, C. E., Suci, G. J., Tannenbaum, P. H. (1957), *The measurement of meaning*, University of Illinois Press, Urbana, USA.
- Pavlenko A. (2005), *Emotions and Multilingualism*, New York: Cambridge University Press.
- Pell, M.D., Monetta, L., Paulmann, S. & Kotz, S.A. (2009), Recognizing emotions in a foreign language, in: *Journal of Nonverbal Behavior*, 33: 107-120.
- Picard, R. (1997), *Affective Computing*, Cambridge: MIT Press.
- Piot, O. (1999), Experimental study of the expression of emotions and attitudes in four languages, in: *Proceedings of the 14th International Conference of Phonetic Sciences (ICPhS)*, San Francisco, USA, pp. 369–370.
- Plutchik, R. (1980a), *Emotion: A Psycho-Evolutionary Synthesis*. New York: Harper & Row.
- Poggi I., Magno Caldognetto E. (2004), Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici, in: Albano Leoni F., Cutugno F., Pettorino M., Savy R. (a cura di), *Atti del Convegno "Italiano parlato"* (Napoli 14-15/2/2003), Napoli: D'Auria Editore, CD-Rom.
- Poggi, I. (2008), *La mente del cuore. Scienze cognitive ed emozioni*, Roma: Armando Editore.
- Rintell, E. (1984), "But how did you FEEL about that?": The learner's perception of emotion in speech, *Applied Linguistics* 5(3): 255-264.
- Romito, L. (2000), *Manuale di Fonetica articolatoria, acustica e forense*, Centro Editoriale e Librario, Università della Calabria, Rende.
- Romito, L., Galatà, V., Lio, R. (2006), Fluency Articulation and Speech Rate as new parameters in the Speaker Recognition, in: *Atti del convegno "III Congreso de Fonética Experimental"*, Santiago de Compostela, 26-24 ottobre, 2005, Xunta de Galicia, pp. 537-549.

- Roseberry-McKibbin, C., Brice, A. (1999), The perception of vocal cues of emotion by Spanish-speaking limited English proficient children, *Journal of Children's Communication Development*, 20(2): 19-25.
- Rottenberg, J., Ray, R. D., & Gross, J. J. (2007), Emotion elicitation using films, in: J. A. Coan & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment*, London: Oxford University Press, pp. 9-28.
- Salovey, P. & Mayer, J. (1990), Emotional Intelligence, in: *Imagination, Cognition and Personality*, 9(3): 185-211.
- Sawamura, K., Dang, J., Akagi, M., Erickson, D. *et al.* (2007), Common factors in emotion perception among different cultures, in: *Proceedings of ICPHS 2007*, Saarbrücken, Germany, pp. 2113-2116.
- Scherer, K. R. (1984), On the nature and function of emotion: a component process approach, in: Klaus R. Scherer and Paul Ekman (Eds.), *Approaches to emotion*, Hillsdale, NJ: Erlbaum, pp. 293-317.
- Scherer, K. R. (1986), Vocal affect expression: A review and a model for future research, *Psychological Bulletin*, 99: 143-165.
- Scherer, K. R. (1988), Criteria for Emotion-Antecedent Appraisal: A Review, in: V. Hamilton, G.H. Bower and N.H. Frijda (Eds.), *Cognitive Perspectives on Emotion and Motivation*, Dordrecht: Kluwer, pp. 89-126.
- Scherer, K. R. (1988), *Facets of emotion: Recent research*, Hillsdale, NJ: Erlbaum.
- Scherer, K. R. (1989), Vocal correlates of emotional arousal and affective disturbance, in H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology*, New York: Wiley, pp. 165-197.
- Scherer, K. R. (1994), Affect Bursts, in: S.H.M. van Goozen, N.E. VandePoll, & J. A. Sergeant (Eds.), *Emotions: Essays on Emotion Theory*, Hillsdale, NJ: Erlbaum, pp. 161-193.
- Scherer, K. R. (2000), A Cross-cultural Investigation of Emotion Inferences from Voice and Speech: Implication for Speech Technology, *Proceedings of the ICSLP 2000*, 2: 379-382.
- Scherer, K. R. (2001), The future of emotion: foreword to the special issue, *Social Science Information*, 40(1): 5-10.

- Scherer, K. R. (2003), Vocal Communication of Emotion: a Review of Research Paradigm, *Speech Communication*, 40: 227-256.
- Scherer, K. R., Banse, R., Wallbott, H.G. (2001), Emotion Inferences from Vocal Expression Correlate across Languages and Cultures, *Journal of Cross-Cultural Psychology* 32(1): 76-92.
- Scherer, K. R., Grandjean, D., Johnstone, L. T., G. Klasmeyer, T. B. (2002), Acoustic correlates of task load and stress, in: *Proceedings of the International Conference on Spoken Language Processing (ICSLP '02)*, 3: 2017-2020.
- Scherer, K. R., Johnstone, T. & Klasmeyer, G. (2003), Vocal expression of emotion, in: Davidson D. J., Goldsmith, H. & Scherer, K. R. (a cura di), *Handbook of the Affective Sciences*, Oxford University Press, New York and Oxford, pp. 433-456.
- Scherer, K. R., Ladd, D. R. & Silverman, K. E. A. (1984), Vocal cues to speaker affect: Testing two models, *Journal of the Acoustical Society of America*, 76: 1346-1356.
- Scherer, K. R., Wranik, T., Sangsue, J., Tran, V. and Scherer, U. (2004), Emotions in everyday life: probability of occurrence, risk factors, appraisal and reaction patterns, *Social Science Information*, 43(4): 499-570.
- Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T. (1991), Vocal cues in emotion encoding and decoding, *Motivation and Emotion*, 15: 123-148.
- Schröder, M. (2004), Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions, *Proc. Workshop on Affective Dialogue Systems: Lecture Notes in Computer Science*, Kloster Irsee, Germany, 2004, pp. 209-220.
- Schröder, M. (2004), Speech and Emotion Research: An overview of research frameworks and a dimensional approach to emotional speech synthesis, *PhD thesis, PHONUS 7, Research Report of the Institute of Phonetics*, Saarland University.
- Seibert, P. S., Ellis, H. C. (1991), A convenient self-referencing mood induction procedure, *Bulletin of the Psychonomic Society*, 29: 121-124.
- Shahid, S., Krahmer, E., Swerts, M. (2008), Real vs. acted emotional speech: comparing South-asian and Caucasian speakers and observers, in *Speech Prosody-2008*, pp. 669-672.

- Shigeno, S. (1998), Cultural similarities and differences in the recognition of audio-visual stimuli, *5th International Conference on Spoken Language Processing*, paper 1057, pp. 149-152.
- Shigeno, S. (2003), Recognition of emotion transmitted by vocal and facial expression: Comparison between the Japanese and the American, in: *The AGU Journal of Psychology*, 3: 1-8.
- Shochi, T., Aubergé, V., Rilliard, A. (2007), Cross-Listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes, in: *Proceedings of ICPHS 2007*, Saarbrücken, Germany, pp. 2097-2100.
- Solomon, D., Ali, F. A. (1975), Influence of verbal content and intonation on meaning attributions of first-and-second language speakers, *Journal of Social Psychology*, 95: 3-8.
- Solomon, R. C. (1993), The Philosophy of Emotions, in: Lewis, M., Haviland, J. M. (Eds.), *Handbook of Emotions*, Guilford Press, New York, pp. 3-15.
- Sutherland, G., Newman, B., Rachman, S. (1982), Experimental investigations of the relations between mood and intensive unwanted cognition, *British Journal of Medical Psychology*, 55: 127-138.
- Thompson, W. F. & Balkwill, L. L. (2006), Decoding speech prosody in five languages, *Semiotica*, 158: 407-424.
- Tisato, G., Cosi, P., Drioli, C., Tesser, F. (2005), INTERFACE: a New Tool for Building Emotive/Expressive Talking Heads, in *Proc. INTERSPEECH 2005*, Lisbon, Portugal, pp. 781-784.
- Van Bezooijen, R., Otto, S., Heenan, T. (1983), Recognition of vocal expressions of emotion: a three-nation study to identify universal characteristics, in: *Journal of Cross-Cultural Psychology*, 14(4): 387-406.
- Van Brakel, J. (1993), Emotions: a cross-cultural perspective on forms of life, in W.M. Wentworth and J. Ryan (eds.), *Social Perspectives on Emotion*, Greenwich USA: JAI Press, vol. 2: 179-237.
- Velten, E. (1968), A laboratory task for induction of mood states, *Behaviour Research and Therapy*, 6: 473-482.

- Verdolini, K. (1994), Voice disorders, in: J. Bruce Tomblin, H. L. Morris, D. C. Spriestersbach (Eds.), *Diagnosis in speech-language pathology*, San Diego : Singular Publishing Group, pp. 247-306.
- Ververidis, D. & Kotropulos, C. (2006), Emotional speech recognition: Resources, features, and methods, *Speech Communication*, 48: 1162-1181.
- Vidrascu, L., Devillers, L. (2005), Detection of real-life emotions in call centers, in *INTERSPEECH-2005*, pp. 1841-1844.
- Wallbott, H. G. & Scherer, K. R. (1986), Cues and channels in emotion recognition, in: *Journal of personality and social psychology*, 51: 690-699.
- Westermann, R., Spies, K., Stahl, G., Hesse, F.W. (1996), Relative effectiveness and validity of mood induction procedures: a meta-analysis, in: *European Journal of Social Psychology*, 26(4): 557-580.
- Williams, C. E., & Stevens, K. N. (1972), Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, 52: 1238-1250.
- Wilting, J., Krahmer, E., Swerts, M. (2006), Real vs. acted emotional speech, in *INTERSPEECH-2006*, pp. 805-808.
- Zajonc, R. B. (1980), Feeling and Thinking: Preferences Need No Inferences, *American Psychologist*, 35(2): 151-175.
- Zajonc, R. B. (1984), On the primacy of affect, *American Psychologist*, 39(2): 117-123.

Sitografia

<http://emotion-research.net/wiki/Databases> (data consultazione: 10 dicembre 2007)

<http://www.elra.info/>

<http://www.langsci.ucl.ac.uk/ipa/index.html> (data consultazione: 13 settembre 2008).

<http://www.psychology.org/links/Resources/Software/> (data consultazione: 24 febbraio 2009).

<http://www.questionmark.com/us/perception/> (data consultazione: 2 marzo 2009)

<http://www.treccani.it/Portale/homePage.html> (data consultazione: 14 marzo 2008)

<http://hyperphysics.phy-astr.gsu.edu/HBASE/music/voicecon.html#c1> (data consultazione: 16 novembre 2007)

<http://hyperphysics.phy-astr.gsu.edu/HBASE/hframe.html> (data consultazione: 16 novembre 2007)

<http://hyperphysics.phy-astr.gsu.edu/HBASE/music/fant.html> (data consultazione 18 novembre 2007)

<http://plato.stanford.edu/entries/james/> (data consultazione: 18 aprile 2008)

<http://www.yorku.ca/earmstro/journey/larynx.html>

Softwares and tools

Bresin, R. & Lindberg, J. (2008), *SKATTA 1.1-beta5: Multiplatform system for the design and running of audio and visual perception tests*, [Computer program], retrieved from <http://sourceforge.net/projects/skatta/>.

Boersma, P. & Weenink, D. (2009), *Praat: doing phonetics by computer* [Computer program], retrieved from <http://www.praat.org/>.

SPSS 16.0 for Windows, Rel. 16.0.1, 2007, Chicago: SPSS Inc.

Questionmark™ Perception™ version 4.3, <http://www.questionmark.com>, in licenza al
Dipartimento di Linguistica - Università della Calabria.

Appendice I - Parametri acustici e psicoacustici più comuni

Table 23.1. Overview of Major Acoustic-Phonetic and Psychoacoustic Parameters

1. Acoustic-phonetic parameters

Speech rate	Number of speech segments per time unit
F0	Fundamental frequency (vibration rate of vocal folds)
F0 perturbation	Slight variations of duration of glottal cycle
F0 contour	Fundamental frequency values plotted over time (intonation or "speech melody")
Intensity	Average squared amplitude within a predefined time segment ("energy")
Spectral energy distribution	Relative amount of energy within predefined frequency bands
Spectral slope	Linear regression of energy distribution in the frequency band above 1 kHz
Laryngalization	Sudden change of oscillation mode of vocal folds (usually to double glottal cycle duration)
Tremor	Regular modulation of glottal cycle duration
Jitter	Regular or irregular variation of glottal cycle duration
Shimmer	Regular or irregular variation of amplitude maxima in subsequent glottal cycles
HNR	("Harmonic-to-noise-ratio") Ratio between harmonic and aperiodic signal energy
GNE	"Glottal-noise-excitation," a measure of hoarseness
Inverse filtered glottal pulse	Transglottal airflow estimated by inverse filtering techniques
Formant	(Time-varying) resonance of vocal tract (significant energy concentration in the spectrum)
F1	First formant—important for vowel identification
F2	Second formant—important for vowel identification
Formant bandwidth	Width of the spectral band containing significant formant energy (- 3 dB threshold)
Formant precision	Degree to which formant frequencies attain values prescribed by phonological system of a language

2. Psychoacoustic parameters

Perceived loudness	Calculated from weighted energy distributions in specific frequency bands
Perceived pitch	Calculated from the F0 contour taking into consideration the linear and differential glissando-threshold
Perceived rhythm	Rhythmic events calculated from the perceived loudness contour

Note: In many cases, these parameters are aggregated over segments of speech by using measures of central tendency and distribution statistics such as mode, mean, median, variance, standard deviation, and upper or lower 5(1)% of the distribution.

Tabella 43: I parametri acustici e psicoacustici più comuni, tratto da Scherer, Johnstone, Klasmeyer (2003: 438)

Appendice II - Istruzioni fornite agli encoder

Modalità di raccolta A³³¹

Leggi con senso, e in modo naturale e spontaneo, il testo proposto dopo averne preso visione. Puoi ripetere la lettura se non sei soddisfatto della tua performance.

Modalità di raccolta B

Rivedi il testo precedentemente proposto in funzione della spiegazione e del chiarimento forniti. Rileggi con senso, e in modo naturale e spontaneo, i testi tenendo conto dell'emozione desiderata: fatti coinvolgere personalmente dalla situazione e renditi partecipe della storia proposta nel testo (soprattutto nelle parti in cui è prevista la forma dialogica). Prima di passare al brano successivo hai facoltà di ripetere se sei insoddisfatto della tua performance.

Testo 1 - rabbia (collera calda); Testo 2 - disgusto; Testo 3 - gioia; Testo 4 - paura; Testo 5 - sorpresa; Testo 6 - tristezza

Modalità di raccolta C

Ripeti la frase “Non è possibile. Non ci posso credere.” in funzione dell'emozione e dello scenario proposto in ciascun dei testi precedentemente letti.

Modalità di raccolta frase neutra

Leggi le frasi contenute nel blocco di fogli che ti è stato consegnato. Se non sei soddisfatto della tua performance puoi ripetere la frase prima di passare al foglio successivo.

I cetacei vivono in mare.

La settimana è fatta di sette giorni.

Non è possibile. Non ci posso credere.

Tutti gli esseri viventi nascono, crescono, si riproducono e muoiono.

Il grano cresce nei campi.

³³¹ Come già detto, prima e durante le sessioni di registrazione, tutte le istruzioni sono state fornite a ciascun soggetto nella propria lingua madre in forma sia scritta che orale.

Appendice III - Scenari forniti agli *encoder* italiani

Testo 1 - rabbia (collera calda)

È agosto, sono le 8:00 e fuori fa molto caldo. Arrivato in ufficio, sulla tua scrivania trovi una lettera. Il capo ha respinto la tua richiesta di ferie. La giornata inizia male. Visibilmente seccato, stai per sederti quando squilla il cellulare: è una tua amica. Ha anticipato la partenza e arriva alle 14:30 in stazione. Devi andare a prenderla. Non è un problema, tanto finisci alle 14:00. Prima di chiudere la conversazione lei ti prega di non fare tardi. Il tempo vola. È ora di andare. Saluti e te ne vai.

Uscendo dall'ufficio, noti che il tempo è cambiato. Prendi la macchina e ti avvii verso la stazione. Improvvisamente la macchina inizia a traballare. Ti fermi subito, scendi e ti accorgi che hai bucato.

- "Uffa, questa non ci voleva!"

Ti togli la giacca, apri il bagagliaio, tiri fuori il cric e smonti la ruota. Inizia a cadere qualche goccia d'acqua e in lontananza si sente tuonare.

- "Accidenti! Ci mancava anche il temporale! e come se non bastasse arriverò tardi alla stazione."

Mentre sollevi la ruota bucata ti si strappa il pantalone.

- "Maledizione! E ora come faccio ad andare in giro conciato così!"

Ti guardi intorno, guardi l'orologio e ti accorgi che è tardi. La tua amica ti aspetta! Tiri fuori la ruota di scorta e la monti. In tutta fretta togli il cric e la macchina va giù. La ruota di scorta è sgonfia! Innervosito dai un calcio alla ruota ed inizi a perdere la pazienza.

- "Ecco, lo sapevo. Non mi dovevo fidare del gommista. Gliel'avevo detto di controllare anche la ruota di scorta!"

Mentre pensi alla tua amica, che ti aveva pregato di non fare tardi, inizia a piovere a dirotto. Esplosi perdendo il controllo:

- "Merda! Cosa ho fatto di male! Porca miseria! Non è possibile. Non ci posso credere! Perché oggi mi va tutto storto. Prima le ferie, poi la ruota e ora anche la pioggia! E Vanessa che mi aspetta alla stazione!"

Testo 2 - disgusto

Sei appena rientrato da una lunga giornata di lavoro. È tardi e sei stanco. Vai in cucina, apri il frigo: è vuoto. Non hai avuto tempo per fare la spesa.

Sei affamato, perciò decidi di uscire. Cerchi un ristorante vicino casa dove poter mangiare qualcosa. I ristoranti che conosci sono chiusi, ma finalmente ne trovi uno ed entri.

Noti che il ristorante è quasi vuoto, ma tu sei stanco, affamato e non hai altra scelta. Ti siedi. Dopo aver sfogliato il menu decidi di ordinare un ricco antipasto e un appetitoso secondo con contorno.

C'è da attendere un po'. Nel frattempo ti guardi intorno e ti accorgi che il locale è un po' sporco, ma pensi che, dopo tutto, sia normale:

- "Mah, è quasi orario di chiusura. Ci sarà stata sicuramente tanta gente."

Finalmente arriva il cameriere con quello che avevi ordinato. Hai l'acquolina in bocca e non vedi l'ora di mangiare. Guardi nel piatto e pensi:

- "Mmmmm! Quante cose deliziose, da dove inizio?"

Senza perdere altro tempo, prendi la prima cosa che ti capita e te la porti alla bocca. Ma, mentre stai per mordere il boccone, vedi nel piatto un qualcosa che si muove. Sorpreso, lasci cadere la forchetta e fai un balzo all'indietro! Ti avvicini al piatto e noti un grosso verme nero e viscido che striscia. Ti guardi intorno e, visibilmente disgustato, dici:

- "Che schifo! Non è possibile! Non ci posso credere! Perché nel mio piatto? Proprio adesso! Proprio ora che avevo iniziato a mangiare con tanto appetito."

Testo 3 - gioia

Giocando alla lotteria, hai appena scoperto di aver vinto una piccola somma di denaro.

- "Evviva! Che bello!"

Non sono molti soldi, ma finalmente qualcuno dei tuoi tanti sogni si potrà avverare.

Passeggi serenamente nel parco ripensando alla vincita e a come spendere i soldi vinti quando, all'improvviso, senti qualcuno urlare il tuo nome come un matto.

- "Ma chi può essere questo deficiente che urla così in mezzo alla gente? Ma chi è? Che vuole?"

Ti giri incuriosito e ti vedi correre incontro una persona. Man mano che si avvicina lo riconosci. Santo cielo! Non lo vedi da tantissimi anni ma lui non è cambiato per niente. È sempre lo stesso. Non hai più sue notizie da quando eravate a scuola. Resti incredulo, eppure non ti puoi sbagliare, è proprio lui! È una giornata fortunata e sei molto felice: prima la vincita e poi l'inaspettato incontro con un tuo vecchio amico.

Abbracciandolo affettuosamente e con immensa gioia:

- "Antonio, amico mio! Che gioia! Non è possibile. Non ci posso credere! E chi mai pensava ad un incontro del genere ora! Deve proprio essere la mia giornata fortunata! Ma che ci fai qui? Da quanto tempo... Vieni andiamo a farci un aperitivo insieme, così mi racconti cosa hai fatto in questi anni."

Chiacchierando e ridendo allegramente vi allontanate ricordando i vecchi tempi.

Testo 4 - paura

Da qualche giorno non riesci a fare a meno di pensare al biglietto anonimo pieno di minacce ritrovato sotto la porta del tuo negozio. Non può che trattarsi di uno stupido scherzo!

Sei un tipo tranquillo, non hai mai dato fastidio a nessuno, non hai mai avuto diverbi con nessuno. Con la gente del quartiere hai un buon rapporto. E che dire dei vicini, fossero tutti

come loro. Eppure qualcuno quel biglietto deve averlo scritto. Ma chi? Con quali intenzioni? Per quale motivo?

Una sera, come al solito ad un'ora tarda, mentre chiudi la serranda del negozio, tutto diventa terribilmente chiaro. La situazione diventa surreale. Al di là della strada, alle tue spalle, tre noti tipacci, giovani, con la testa rasata e la giacca di pelle nera. Dagli sguardi che si scambiano capisci che aspettano proprio te.

- "Dio mio, che facce! E ora?"

Ti senti gelare, incapace di reagire. Guardi a destra e a sinistra in cerca di qualcuno, ma la strada è deserta e poco illuminata. Ti senti preso in trappola. Resti immobile con le spalle alla serranda!

I tre giovani si guardano intorno e vengono verso di te. Hanno delle spranghe in mano. Vuoi fuggire, ma la paura è tanta e non riesci a muovere un dito. Uno dei tre allunga il passo, ed eccolo davanti a te. Gli altri lo raggiungono. Lo guardi negli occhi, tenti di parlare:

- "Non capisco cosa volete. Io non ho fatto nulla di male. Così all'improvviso ... Ma cosa volete?"

Con la coda dell'occhio noti che uno di loro ha in mano un coltello affilato. Ti senti un brivido giù per la schiena. La paura ti assale:

- "Io non so niente, cos'è tutta questa storia! Non è possibile. Non ci posso credere! Ma cosa volete farmi!!! Oh Dio mio è incredibile, vi giuro che io ... ma cosa fate, fermi. Aiuto! Io non capisco, aiuto!"

Testo 5 - sorpresa

Barbara, una tua carissima amica, ti ha appena chiamato al telefono. Vuole incontrarti, ha bisogno di vederti per parlare di una questione che riguarda Tom, il suo ragazzo. Sei in biblioteca all'università, perciò vi date appuntamento al bar che sta proprio di fronte alla biblioteca, dove lei ti raggiungerà tra un'ora.

Nel frattempo pensi a cosa possa essere successo. Non vi vedete da un po' di tempo ma conosci benissimo Tom.

- "Cosa mai può essere successo? Che avrà combinato?"

L'ultima volta che vi siete visti, Tom ti aveva detto che, Barbara e lui, avevano deciso di sposarsi.

Si è fatta ora e, come d'accordo, raccogli le tue cose ed esci dalla biblioteca per raggiungere il bar. Barbara è già lì. Vi salutate e lei inizia a raccontarti del viaggio che lei e Tom hanno appena fatto a Londra. Sembra un fiume in piena. È serena, felice. Tu segui con attenzione il suo racconto quando, ad un certo punto, Barbara si interrompe e pronuncia la frase: "Mi ha lasciata."

Non riuscendo a capire, la guardi negli occhi e chiedi spiegazioni:

- "Che vuol dire mi ha lasciata? Perché?"

Barbara sta zitta, poi scoppia a ridere. Inizialmente pensi che lei ti stia facendo uno dei suoi soliti scherzi ma, improvvisamente, smette di ridere e, con lo sguardo perso e serio, dice: "Mi ha lasciata per un altro ragazzo. È gay."

Tu convinto che stia scherzando:

- "Ma cosa dici? Mi stai prendendo in giro, vero?"

Barbara, però, conferma che non si tratta affatto di uno scherzo, anzi... Ti racconta come sono andate le cose e tu ti convinchi che la situazione è seria. Non avresti mai potuto immaginare una cosa del genere e la notizia ti lascia a dir poco sorpreso:

- “Ma sei sicura? Forse ti sbagli. Non è possibile. Non ci posso credere! Proprio ora che avevate deciso di sposarvi.”

Cerchi di consolare Barbara, ma non sai che dire.

Testo 6 - tristezza

Non riesci proprio a dimenticarla.

Tutte le mattine ti alzi e, quando ti guardi allo specchio, cerchi di convincerti che la vita continua. Ti prepari e ti sistemi per andare a lavoro. Ma che senso ha. La persona che ami e che tutte le mattine si sveglia al tuo fianco non c'è più. Ludovica se n'era andata. Convincendoti che sia tutto un brutto sogno, vai di là in cucina, pensando di trovarla lì a preparare la colazione, di fretta, come al solito. Ma lei non c'è.

Improvvisamente ti ritrovi in lacrime, seduto per terra, a pensare ai bei momenti passati insieme. Ripensi alle cose più banali ed irrilevanti che, improvvisamente, sembrano aver acquistato un valore indescrivibile.

Squilla il telefono e parte la segreteria telefonica: è la sua voce, ma lei non c'è più. Gridi il suo nome che riecheggia nella casa vuota.

- “Perché te ne sei andata, perché mi hai lasciato? Come farò senza di te.”

Lei però non ti potrà mai più rispondere. Ripensi a quando vi siete conosciuti. Tu le parlavi e lei sembrava rapita dalle tue parole. Aveva lo sguardo perso, era innamorata! Eravate innamorati! L'uno viveva per l'altra. La vedi dappertutto e in ogni cosa che ti circonda. Ovunque percepisci il suo odore. La tristezza ti assale:

- “Non è giusto! Perché proprio a noi due. Non è possibile. Non ci posso credere! Ho bisogno di te. E oggi ho anche la riunione con i giapponesi! Così non può andare avanti.”

Appendice IV - Scenari forniti agli encoder francesi

Texte 1 - colère chaude

C'est le mois d'août. Il est 8 heures et il fait déjà très chaud. Arrivé au bureau, tu trouves une lettre sur ta table de travail. Ton chef a repoussé ta demande de congé. La journée commence mal. Visiblement fâché, tu vas t'asseoir. Ton portable retentit: il s'agit d'une de tes amies. Elle a anticipé son départ, elle arrive à la gare à quatorze heures trente. Tu dois aller la chercher. Ce n'est pas un problème car tu quittes le travail à quatorze heures. Elle te prie de ne pas arriver en retard. Le temps file. Il faut y aller. Tu dis au revoir et tu pars.

En sortant du bureau tu remarques que le temps commence à changer. Tu prends ta voiture et (tu) te diriges vers la gare, mais la voiture commence à cahoter. Tu t'arrêtes immédiatement, (tu) descends et (tu) t'aperçois que tu as crevé.

- « Zut alors ! il ne manquait que ça! »

Tu retires ta veste, (tu) ouvres le coffre, (tu) prends le cric et (tu) commences les opérations de changement du pneu. Il tombe quelques gouttes d'eau et on entend tonner au loin.

- « Oh là là ! Il va y avoir un orage! Et... je vais arriver en retard à la gare. »

Pendant que tu soulèves le pneu crevé, ton pantalon se déchire.

- « Ça alors ! Ça va faire une belle vue ! »

Tu regardes autour de toi, (tu) regardes aussi ta montre et (tu) te rends compte qu'il est tard. Ton amie t'attend! Tu sors la roue de secours et (tu) la montes. Tu enlèves le cric à toute vitesse et la voiture descend. Mais, manque de pot, la roue de secours aussi est dégonflée! Enervé, tu envoies un coup de pied à la roue et tu commences sérieusement à te fâcher.

- « Qu'est-ce que je suis bête! Je (ne) devais pas faire confiance à mon garagiste. Je lui avais dit de contrôler la roue de secours! »

Pendant que tu penses à ton amie qui t'avait prié de ne pas arriver en retard, il commence à pleuvoir à verse. Tu perds le contrôle et tu explotes:

- « Putain de merde ! Ça fait mal ! Oh là là ! C'est incroyable! Aujourd'hui tout fout le camp. D'abord mon congé, ensuite le pneu et maintenant le déluge! Et Vanessa qui m'attend à la gare! »

Texte 2 - dégoût

Tu viens de rentrer après une longue journée de travail. Il est tard et tu es fatigué. Tu vas à la cuisine, tu ouvres le frigo mais il est vide. Tu n'as pas eu le temps de faire tes courses. Tu as faim donc tu décides de sortir. Tu cherches un restaurant près de chez toi. Les restaurants que tu fréquentes sont tous fermés. Enfin tu trouves un bistrot ouvert et tu entres. Tu remarques

que la salle est presque vide mais tu es fatigué, tu as faim et tu n'as pas d'autres choix. Tu t'assois. Après avoir jeté un coup d'œil à la carte, tu décides de commander des crudités en entrée et un bon steak-frites. Il faut attendre un peu. Tu regardes autour de toi et (tu) t'aperçois que la pièce est un peu sale, mais tu te dis:

- « C'est peut-être parce qu'il y a eu beaucoup de monde : ils vont fermer sous peu. »

Enfin ton entrée arrive. Tu as vraiment faim et tu commences par les carottes. Mais pendant que tu manges tes carottes, tu vois quelque chose qui bouge dans l'assiette. Étonné, tu regardes mieux et tu vois un petit ver qui se promène dans le céleri. Tu laisses tomber la fourchette, tu regardes autour de toi et, visiblement dégoûté, tu dis:

- « Quelle horreur! Oh là là ! C'est incroyable! Pas maintenant! Juste au moment où j'avais commencé à manger ! Et dire que j'avais faim... »

Texte 3 - joie

Tu as joué au loto et tu as gagné un peu d'argent.

- « Formidable ! Extra ! »

Il ne s'agit pas d'une grosse somme, mais tu pourras enfin réaliser quelqu'un de tes rêves. Tu te promènes tranquillement dans le parc et tu penses à comment tu pourras dépenser l'argent gagné. Tout à coup tu entends quelqu'un qui hurle ton nom.

- « C'est qui cet idiot qui crie comme un veau ? Qu'est-ce qu'il veut? »

Tu te retournes et tu vois un type qui cours vers toi. Au fur et à mesure qu'il se rapproche tu le reconnais. C'est le vieux Lulu ! Ça faisait vraiment longtemps que tu ne le voyais pas, probablement depuis le collège, mais lui, il n'a pas changé. Tu hésites, mais tu ne peux pas te tromper : c'est bien lui! C'est vraiment ta journée et tu es très heureux: d'abord le loto et ensuite cette rencontre inattendue...

Tu l'embrasses chaleureusement, avec une joie immense:

- « C'est toi, mon vieux Lulu ? C'est bien toi ? Quelle joie! Oh là là ! C'est incroyable! Je n'aurais jamais imaginé te rencontrer ici... C'est vraiment ma journée! Mais que qu'est-ce que tu fais ici? Depuis combien de temps... Allez, on va prendre un pot et tu me raconteras ce que tu as fait pendant tout ce temps. »

Ainsi, en bavardant et en riant vous commencez à vous rappeler le bon vieux temps...

Texte 4 - peur

Depuis quelques jours tu ne peux pas éviter de penser au billet anonyme de menaces retrouvé sous la porte de ton magasin. Il doit s'agir d'une plaisanterie stupide!

Tu es un type tranquille, tu n'as jamais importuné personne, tu n'as jamais eu de disputes avec personne. Tu as un bon rapport avec les gens du quartier. Et tu n'as pas à te plaindre de tes voisins : tout le monde devrait être comme eux! Pourtant... quelqu'un doit avoir écrit ce billet. Mais qui? Pourquoi ? Avec quelles intentions?

Un soir, pendant que tu fermes le rideau de ton magasin, tout devient terriblement clair. Au-delà de la rue, tu aperçois trois sales types, jeunes, la tête rasée, portant des vestes en cuir noir. Aux regards qu'ils échangent tu comprends que c'est toi qu'ils attendent.

- « Le voilà, mon billet... Des voyous! Et maintenant, qu'est-ce que je vais faire? »

Tu te sens bloqué, tu es incapable de réagir. Tu regardes à droite et à gauche à la recherche de quelqu'un qui t'aide, mais la rue est déserte et peu éclairée. Tu te sens pris au piège. Tu restes immobile, le dos contre le rideau. Les trois types avancent vers toi. Ils ont des barres de fer. Tu voudrais te sauver, mais tu as tellement peur que tu n'arrives même pas à lever le petit doigt. Et voilà qu'un des trois types est devant toi. Les autres le suivent. Tu le regardes dans les yeux et (tu) essaies de parler:

- « Je ne comprends pas ce que vous voulez. Je n'ai rien fait de mal. Qu'est-ce que vous me voulez ? »

Du coin de l'œil, tu remarques aussi qu'un des trois compères a un couteau à la main. Un frisson de peur secoue ton dos. La peur t'attaque:

- « Je ne comprends pas... Oh là là ! C'est incroyable! Oh mon Dieu... je vous jure que je... mais qu'est-ce que vous faites? Arrêtez! Au secours! Au secours ! »

Texte 5 - surprise

Barbara, ta meilleure amie, vient de te donner un coup de fil. Elle veut te rencontrer, elle a besoin de te voir pour parler d'un problème qui concerne Patrick, son petit ami. Tu es dans la bibliothèque à l'université donc vous vous donnez rendez-vous au café en face de la bibliothèque : elle dit qu'elle arrivera dans une heure. En attendant, tu penses à ce qui peut être arrivé. La dernière fois que tu as vu Patrick, il t'avait dit que Barbara et lui avaient décidé de se marier... Maintenant, c'est l'heure du rendez-vous, tu prends tes affaires et (tu) vas au café. Barbara est déjà là. Vous vous faites la bise et elle commence à te raconter du voyage qu'elle a fait à Londres avec Patrick. Elle est tranquille, heureuse. Tu suis avec attention son récit, mais tout à coup Barbara s'interrompt et dit:

- « Il m'a quittée! »

Tu n'arrives pas à comprendre, tu la regardes dans les yeux et (tu) lui demandes des explications: - « Qu'est-ce que ça veut dire il m'a quittée? Pourquoi? »

Barbara reste silencieuse puis elle éclate de rire. Tu penses qu'elle plaisante mais tout à coup, elle arrête de rire et, le regard perdu et sérieux, (elle) dit:

- « Il m'a quittée pour un mec. Il est omo! »

Tu penses encore qu'elle plaisante:

- « Mais qu'est-ce que tu dis? Tu te moques de moi, ou quoi? »

Mais Barbara dit qu'il ne s'agit pas du tout d'une plaisanterie, ou mieux... Elle te raconte tout ce qui s'est passé et toi, tu comprends que c'est vrai. Tu n'aurais jamais pu imaginer cela et la nouvelle te laisse étonné:

- « Mais... tu es sûre? Peut-être que tu te trompes... Oh là là ! C'est incroyable! Juste au moment où vous aviez décidé de vous marier! »

Tu essaies de consoler Barbara mais tu ne sais que dire...

Texte 6 - tristesse

Tu ne peux vraiment pas l'oublier.

Tous les matins tu te lèves et quand (tu) te regardes dans le miroir, (tu) essaies de te convaincre toi-même que la vie continue. Tu te prépares pour aller au travail. Mais tout cela

n'a pas de sens! La personne que tu aimes et qui se réveille près de toi chaque matin n'existe plus. Dominique est morte. Tu te dis qu'il s'agit simplement d'un cauchemar et tu vas à la cuisine en pensant que tu vas la trouver là-bas en train de préparer le petit déjeuner, comme d'habitude. Mais elle, elle n'existe plus.

Tout à coup tu te retrouves en larmes. Le souvenir des beaux moments passés ensembles afflue et même les détails les plus banales et insignifiants acquièrent une valeur indescriptible. Le téléphone sonne et le répondeur commence: c'est sa voix mais elle, elle n'existe plus. Tu cries son nom qui retentit dans la maison vide.

- « Pourquoi tu es morte? Pourquoi tu m'as quittée? Qu'est-ce que je vais faire sans toi? »

Mais elle, elle ne pourra plus te répondre. Tu revois le moment de votre rencontre. Tu lui parlais et elle, elle semblait ravie par tes mots. Elle avait le regard perdu, elle était amoureuse! Vous étiez amoureux! Vous viviez l'un pour l'autre. Tu la revois partout et dans chaque objet. Son parfum est partout. La tristesse t'attaque:

- « Ce n'est pas juste! Pourquoi à nous deux? Oh là là ! C'est incroyable ! Ça ne peut pas continuer comme ça. J'ai besoin de toi... Et aujourd'hui j'ai aussi la réunion avec les Japonais! »

Appendice V - Scenari forniti agli encoder inglesi

Text 1 - hot anger

It's August. It's 8.00 o'clock and it's very hot outside. When you get to the office, you find a letter on your desk. Your boss has rejected your request for a holiday. The day starts badly. Visibly irritated, you are about to sit down when your mobile rings. It is your girlfriend. She has anticipated her departure and is arriving at 2.30 at the station. You have to go to pick her up. It's no problem – you finish at 2.00 anyway. Before finishing the conversation, she begs you not to be late. The time flies. It's time to go. You say goodbye and leave. As you leave the office, you notice that the weather has changed. You get the car and set off for the station. Suddenly the car starts to wobble. You stop immediately, get out and you realise that you've got a puncture. Oh dear, that's the last thing I needed! You take off your jacket, open the boot, take out the jack and remove the tyre. Drops of rain start to fall and in the distance you hear the rumble of thunder.

- "Damn! All we needed was a thunder storm. As if that wasn't enough, now I'm going to get to the station late".

As you lift the punctured tyre you rip your trousers.

- "Damnation! How can I move around dressed like this?"

You look around, look at your watch and realise that it's late. Your friend is waiting for you! You take out the spare wheel and fit it. Hastily you remove the jack and the car descends. The spare is flat! In rage you kick the wheel and start losing patience.

- "There, I knew it. I shouldn't have trusted that tyre dealer, I told him to check the spare!"

While you are thinking of your friend who had prayed for you not to be late, it started to pour with rain. And you explode, losing control completely:

- "Shit! What did I do wrong. Damn it! It can't be. I cannot believe it! Why has everything gone wrong today? First the holidays, then the wheel and now the rain as well! And Vanessa's waiting for me at the station!"

Text 2 - disgust

You've just got home after a long day at work. It's late and you're tired. You go into the kitchen and open the fridge, but it's empty. You haven't had the time to do the shopping. You're very hungry, so you decide to go out. You look for a restaurant near home to eat something. The ones you know are closed but finally you find one and go in. You notice that the restaurant is almost empty, but you're tired and hungry and you have no other choice. You sit down. After having leafed through the menu you decide to order a big starter and a tasty

main course with a side dish. There's a bit of a wait. Meanwhile, you look around and you notice that the place is a little dirty but you think that, after all, it's normal and you think:

- "Well, it's almost closing time. There must have been a lot of people".

Finally the waiter arrives with your order. Your mouth is watering and you can't wait to start eating. You look at the plate and think:

- "Mmmm! So many delicious things. Where do I start?"

Without wasting time, you pick up the first thing you light upon and put it to your mouth, but as you are about to bite into the morsel, you see something moving on the plate. Surprised, you drop your fork and jump back! You approach the plate and notice a big slimy black worm squirming about. You look around you, and clearly disgusted you exclaim:

- "Yuk! It can't be. I cannot believe it! Why my plate? Now of all times! Now that I started eating with such hunger!"

Text 3 - joy

By playing the lottery, you've just discovered that you've won a small sum of money.

- "Hurray! That's great!"

It's not a lot of money, but finally you can realise one of your dreams. You go for a walk in the park, thinking of how to spend the money you've won when all of a sudden you hear someone calling your name like a maniac.

- "Who is this hothead, shouting like that in front of people? Who is he? What does he want?"

Curious, you turn round and see a person coming towards you. As he gets nearer you recognise him. Heavens! You haven't seen him for years but he hasn't changed a bit. He's still the same. You haven't heard news of him since you were at school. You are amazed, but you can't be mistaken. It's him alright. This is a lucky day and you're very happy: first the win and then this unexpected meeting with your old friend. Embracing him affectionately and with great joy:

- "Antonio, my friend! What a joy! It can't be. I cannot believe it! Who would have thought of meeting like this now! This has got to be my lucky day! But what are you doing here? How long it's been... Come on, let's go have an drink together and you can tell me what you've been doing all these years."

Chatting and laughing you set off remembering old times.

Text 4 - fear

For some days you have been unable not to think of the anonymous note full of threats you found under the door of your shop. It can't be anything other than a stupid joke! You are a tranquil type of person – you've never given anybody any bother – you've never had an argument with anyone. You get on well with the people in your area. And what can you say about the neighbours. If only everyone was like them. Yet... someone must have written that note. But who? And why? For what reason? One evening, as usual at a late hour while you are closing the shutters of the shop, everything becomes terribly clear. The situation becomes surreal. On the other side of the street, behind you, three well known hoodlums, youths with

skin head haircuts and black leather jackets. From the looks they exchange you realise that it's you they're waiting for.

- "Oh God, what faces! And now what?"

You freeze, unable to react. You look right and left for somebody, but the street is empty and poorly lighted. You feel trapped. You stay still with your back to the shutters.

The three youths look about them and come towards you. They are carrying iron bars. You want to run away but you are so afraid you can't move a muscle. One of them steps forward and is right in front of you. The others join him. You look him in the eyes. You try to speak.

- "I don't know what you want. I've done nothing wrong. Just like this, all of a sudden. What do you want?"

Out of the corner of your eye you see that one of them has a sharp knife in his hand. You feel a shiver run down your spine. Fear takes hold of you.

- "I know nothing about any of this! It can't be. I cannot believe it! What are you doing to me? Oh my God. It's incredible, I swear that I...what are you doing? Stop! Help. I don't understand! Help!"

Text 5 - surprise

Barbara, a very dear friend of yours, has just called you on the phone. She wants to meet you, needs to speak to you about something to do with her boyfriend Tom. You are in the university library, so you make an appointment to meet in the bar in front of the library where she will reach you in an hour. Meanwhile, you think about what might have happened. You haven't seen each other for a while but you know Tom very well.

- "Whatever can have happened? What's he got up to?"

The last time you saw him Tom had said that he and Barbara had decided to get married. The time comes, and as agreed, you get your things together and leave the library to go to the bar. Barbara is already there. You greet each other and she starts to tell you about the trip that she and Tom have just made to London. She's full of news, serene, happy. You follow her account attentively when all of a sudden she stops and says, "he's left me".

Unable to understand you look her in the eyes and ask her to explain:

- "What do you mean he's left you? Why?"

Barbara is silent, then she bursts out laughing. At first, you think she's playing one of her usual jokes on you, but suddenly she stops laughing and with a look that is lost and serious, says "He's left me for another guy. He's gay."

Convinced that she's joking, you say:

- "What are you saying? You're pulling my leg, aren't you?"

Barbara, however, confirms that it's not a joke, at all, anything but... She tells you how things have gone and you realise that the situation is serious. You would never have imagined such a thing and the news leaves you surprised to say the least.

- "But are you sure? Maybe you're wrong. It can't be. I cannot believe it! Just when you've decided to get married."

You try to console Barbara, but you don't know what to say...

Text 6 - sadness

You just can't forget her. Every morning you get up and when you look in the mirror you try to convince yourself that life goes on. You prepare and get ready to go to work. But what sense is there in it? The person you love and who wakes beside you every morning, is no longer there. Meg has gone. Persuading yourself that it's all been a horrible dream you go into the kitchen imagining you'll find her there preparing breakfast, in a hurry as usual. But she isn't there. Suddenly you find yourself sitting on the floor crying and think of the good times spent together. You think again of the most banal and irrelevant things which suddenly appear to have acquired an indescribable value. The phone rings and the message recorder activates: it's her voice, but she is no longer there. You shout her name, which echoes around the empty house.

- "Why have you gone? Why have you left me? How will I manage without you?"

But she will never be able to answer you. You think again of when you first met. You talked to her and she seemed entranced by your words. She had a lost look. She was in love! You were both in love! The one lived for the other. You see her everywhere and in everything around you. Your notice her scent everywhere. Sadness assails you:

- "It's not right! Why us? It can't be. I cannot believe it! I can't go on like this. I need you. And today I've got a meeting with the Japanese as well!"

Appendice VI - Scenari forniti agli encoder tedeschi

Text 1 - Ärger

Es ist August, es ist 8:00 Uhr und draußen ist es sehr warm. Bei deiner Ankunft im Büro findest du einen Brief auf deinem Schreibtisch. Der Chef hat deinen Urlaubsantrag abgelehnt. Der Tag fängt schlecht an. Du bist sichtlich verärgert, nun ringt das Telefon: es ist eine deiner Freundinnen. Sie hat ihre Abfahrt vorverlegt und kommt um 14:30 am Bahnhof an. Du sollst sie abholen. Es ist kein Problem, du bist um 14:00 Uhr fertig. Bevor sie auflegt, bittet sie dich nicht zu spät zu kommen. Die Zeit geht schnell um. Es ist Zeit zu gehen. Du grüßt und verlässt das Büro.

Beim verlassen des Büros merkst du, dass sich das Wetter umgestellt hat. Du nimmst das Auto und fährst zum Bahnhof. Plötzlich fängt das Auto an zu schlingern. Du hältst sofort an, steigst aus und merkst dass du eine Panne hast.

- „Mist, ausgerechnet heute!“

Du ziehst deine Jacke aus, machst den Kofferraum auf, nimmst den Wagenheber raus und den kaputten Reifen ab. Einige Regentropfen beginnen herunter zu fallen und von weitem hört man es donnern.

- „Zum Teufel! Das Gewitter auch noch dazu! Und als ob es nicht schon genug wäre, jetzt komme ich auch noch spät am Bahnhof an.“

Du gehst in die Knie, um den Reifen zu wechseln, und dabei zerreißt deine Hose.

- „Zum Teufel noch mal! Wie soll ich jetzt so herumlaufen!“

Du schaust dich um, siehst auf deine Uhr und merkst, dass es spät ist. Deine Freundin wartet auf dich! Du nimmst den Ersatzreifen und montierst ihn an. In aller Eile nimmst du die Winde weg und das Auto kracht herunter. Der Ersatzreifen ist platt! Nervös geworden gibst du dem Ersatzreifen einen Tritt und fängst an, die Geduld zu verlieren.

- „Da, ich wußte es! Ich hätte mich nicht auf den Reifenreparateur verlassen sollen. Ich hatte ihm gesagt, er soll auch den Ersatzreifen kontrollieren!“

Du denkst an deine Freundin, die dich gebeten hatte, nicht zu spät zu kommen. Es fängt an in Strömen zu gießen, du explodierst und verlierst deine Kontrolle:

- „Scheiße! Was hab ich da nur getan? Verdammte Schweinerei! Das ist nicht möglich. Ich kann es nicht glauben! Warum geht heute alles schief? Zuerst das mit den Ferien, dann der Reifen und jetzt auch noch der Regen! Und Vanessa wartet auf mich am Bahnhof!“

Text 2 - Ekel

Du bist nach einem langen Arbeitstag wieder zu Hause. Es ist spät und du bist müde. Du gehst in die Küche, machst den Kühlschrank auf aber er ist leer. Du hast keine Zeit gehabt um einzukaufen. Du bist hungrig und deshalb entscheidest du dich auszugehen. Du suchst ein Restaurant in der Nähe deiner Wohnung, wo du etwas essen kannst. Die Restaurants, die du kennst sind schon zu, aber endlich findest du eines und trittst ein. Du merkst, dass das Restaurant fast leer ist, aber du bist müde, hungrig und hast keine andere Wahl. Du nimmst Platz. Nachdem du die Speisekarte durchgeblättert hast, entscheidest du eine reichhaltige Vorspeise und ein Fleischgericht mit Beilage zu bestellen. Man muss ein bisschen warten. Inzwischen schaust du dich um und merkst, dass das Restaurant ein bisschen schmutzig ist aber du denkst, dass es eigentlich normal ist:

- „Na ja, es ist bald Zeit zu schließen. Es gab sicher eine Menge Leute hier.“

Endlich kommt der Kellner mit deiner Bestellung. Du hast Speichel im Mund und kannst es nicht erwarten zu essen. Du schaust auf den Teller und denkst:

- „Mmmmm! Wie viele köstliche Sachen, wo fange ich an?“

Ohne weitere Zeit zu verlieren, nimmst du das erste das auf dem Teller ist und fängst an zu essen. Aber während du im Begriff bist, in den Happen zu beißen, siehst du etwas im Teller, das sich bewegt. Überrascht läßt du die Gabel fallen und machst einen Sprung auf dem Stuhl nach hinten! Du näherst dich wieder dem Teller und bemerkst einen großen schwarzen schleimigen Wurm, der dort kriecht. Du schaust dich um und sichtlich angeekelt sagst du:

- „Ist das ekelig! Das ist nicht möglich. Ich kann es nicht glauben! Warum auf meinem Teller? Gerade jetzt! Gerade jetzt wo ich mit so großem Appetit angefangen hatte zu essen.“

Text 3 - Freude

Du hast Lotterie gespielt und du hast gerade entdeckt, dass du eine kleine Geldsumme gewonnen hast.

- “Hurra! Fantastisch!”

Es ist nicht viel Geld, aber endlich kann einer deiner vielen Träume wahr werden. Du spazierst ruhig im Park und denkst an deinen Gewinn und wie du das Geld ausgeben wirst, aber plötzlich hörst du jemand, der wie ein Verrückter deinen Namen schreit.

- „Wer ist denn dieser undisziplinierte Mensch, der inmitten der Leute so schreit! Wer ist er? Was will er?“

Neugierig drehst du dich um und siehst dir jemand entgegen laufen. Sobald er näher kommt, erkennst du ihn. Du lieber Himmel! Du hast ihn seit Jahren nicht mehr gesehen, aber er hat sich überhaupt nicht verändert. Er sieht immer noch gleich aus. Du hast seit den Schuljahren nichts mehr von ihm gehört. Du bist sprachlos, jedoch kannst du dich nicht irren, tatsächlich, er ist es! Es ist ein glücklicher Tag und du bist sehr froh: zuerst der Gewinn und dann das unerwartete Treffen mit einem alten Freund von dir. Du umarmst ihn warmherzig und mit großer Freude:

- „Antonio, mein Freund! Was für eine Freude! Das ist nicht möglich. Ich kann es nicht glauben! Wer hätte je an eine solche Begegnung gedacht! Es muß tatsächlich mein Glückstag

sein! Was machst du eigentlich hier? Wie lange ist das her? Komm, wir gehen zusammen einen trinken, so dass du mir erzählen kannst, was du in den letzten Jahren gemacht hast.“
Fröhlich redend und lächelnd geht ihr weg in Erinnerung an die Vergangenheit.

Text 4 - Angst

Seit einigen Tagen denkst du immer an den Zettel voller Drohungen, den du unter der Tür deines Ladens gefunden hast. Es kann sich nur um einen Scherz handeln!

Du bist ein ruhiger Typ, du hast nie jemandem belästigt, du hast nie mit jemandem einen Wortstreit gehabt. Mit den Leuten aus deinem Stadtviertel hast du gute Beziehungen. Was soll man über die Nachbarn sagen, wären doch nur alle so wie sie! Trotzdem... jemand muss den Zettel geschrieben haben. Aber wer? Mit welcher Absicht? Warum?

Eines Abends, wie üblich spät am Abend, beim Schließen der Rolladen von deinem Laden wird alles klar. Die Situation wird surrealistisch. Auf der Straße gegenüber, hinter dir, sind drei bekannte Sonderlinge, Jungen mit rasiertem Kopf und schwarzer Lederjacke. Von den Blicken, die sie untereinander austauschen, verstehst du, dass sie auf dich warten.

- „Mein Gott, was für Gesichter! Und jetzt?“

Du fühlst dich erstarren, ohne Reaktionsfähigkeit. Du schaust dich nach rechts und links um, auf der Suche nach jemandem aber die Straße ist menschenleer und wenig beleuchtet. Du fühlst dich in einer Falle. Du stehst bewegungslos mit dem Rücken gegen den Rolladen! Die drei Jungen schauen sich um und kommen dir entgegen. Sie haben Eisenstangen in der Hand. Du möchtest abhauen aber die Angst ist groß und du kannst kaum einen Finger bewegen. Einer der drei läuft jetzt schneller und steht jetzt vor dir. Die anderen erreichen ihn. Du siehst ihm in die Augen, du versuchst zu sprechen:

- „Ich verstehe nicht, was ihr wollt. Ich habe nichts Schlechtes getan. So plötzlich... Aber was wollt ihr?“

Verstohlen merkst du, dass einer von ihnen ein scharfes Messer in der Hand hat. Du spürst eine Gänsehaut, die dir den Rücken herunter läuft. Große Angst überkommt dich:

- „Ich verstehe nicht, was soll diese Geschichte! Das ist nicht möglich. Ich kann es nicht glauben! Was wollt ihr mir tun!!! Oh mein Gott ... es ist unglaublich, ich schwöre euch, dass ich.... was macht ihr, halt. Hilfe! Ich verstehe nicht, Hilfe!“

Text 5 - Überraschung

Barbara, eine deiner liebsten Freundinnen, hat dich gerade angerufen. Sie möchte sich mit dir treffen, sie möchte dich sehen, um mit dir über eine Frage, die Tom betrifft, zu sprechen. Du bist in der Bibliothek an der Uni und deshalb macht ihr ab, euch in einer Stunde am Café, das gegenüber der Bibliothek liegt, zu treffen. Inzwischen denkst du daran was passiert sein könnte. Ihr seht euch seit einiger Zeit nicht mehr aber du kennst Tom sehr gut. Du denkst:

- „Was kann eigentlich passiert sein. Was hat er angestellt?“

Das letzte Mal, dass du ihn gesehen hast, hat er dir gesagt, Barbara und er hätten sich entschieden zu heiraten.

Es ist Zeit und, wie abgemacht, sammelst du deine Sachen und verlässt die Bibliothek, um zum Café zu gehen.

Barbara ist schon da. Ihr begrüßt euch und sie fängt sofort an, dir von der Reise nach London, die Tom und sie gerade gemacht haben zu erzählen. Sie ist ein echtes Plappermaul. Sie ist froh, unbeschwert. Du folgst ihrer Erzählung mit Aufmerksamkeit, als Barbara plötzlich in ihrem Redeschwall stoppt und folgenden Satz ausspricht: „Er hat mich sitzen gelassen.“

Du verstehst nicht, du schaust ihr in den Augen und fragst nach Erläuterungen:

- „Was heißt, dass er dich sitzen gelassen hat? Warum?“

Barbara bleibt still, dann bricht sie in Lachen aus. Am Anfang denkst du es handelt sich um einen ihrer gewöhnlichen Scherze aber plötzlich hört sie auf zu lachen und mit ernstem Blick sagt sie: „Er hat mich für einen anderen sitzen gelassen. Er ist Schwul.“

Du glaubst immer noch, es sei ein Scherz:

- „Was sagst du? Du nimmst mich auf die Schippe, nicht wahr?“

Aber Barbara bekräftigt, dass es sich nicht um einen Scherz handelt, im Gegenteil... Sie erzählt dir, wie es gegangen ist und du überzeugst dich, dass die Situation ernst ist. Du hättest nie etwas solcher Art glauben können und die Nachricht überrascht dich etwas:

- „Bist du sicher? Vielleicht irrst du dich. Das ist nicht möglich. Ich kann es nicht glauben! Gerade jetzt da ihr euch entschieden hattet, zu heiraten.“

Du versuchst Barbara zu trösten, aber du weißt nicht was du sagen sollst.

Text 6 - Trauer

Seitdem sie gestorben ist, kannst du sie einfach nicht vergessen.

Jeden Morgen stehst du auf und schaust dich im Spiegel an, im Versuch dich zu überzeugen, dass das Leben weitergeht. Du machst dich fertig um zur Arbeit zu gehen. Welchen Sinn soll das noch haben. Die Person, die du liebst und die jeden Morgen neben dir aufwacht, ist nicht mehr da. Monika ist für immer fort. Überzeugt davon, dass es sich um einen schlechten Traum handelt, gehst du in die Küche mit dem Gedanken, sie beim Frühstückvorbereiten zu finden, in voller Eile, wie immer. Aber sie ist nicht da.

Plötzlich hockst du in Tränen auf dem Boden und denkst an die wunderbaren miteinander verbrachten Momente. Du denkst an die alltäglichsten und unerheblichsten Sachen, die plötzlich einen unbeschreiblichen Wert gewinnen.

Das Telephon klingelt und der Anrufbeantworter schaltet sich ein: es ist ihre Stimme, aber sie ist nicht mehr da. Du schreist ihren Namen, der im leeren Haus widerhallt.

- „Warum bist du weg, warum hast du mich verlassen? Wie soll ich ohne dich weitermachen.“

Sie kann dir aber nie mehr antworten. Du denkst daran, wann ihr euch kennengelernt habt. Du sprachst zu ihr und sie war von deinen Worten wie entführt. Ihr Blick war verloren, sie war verliebt! Ihr wart verliebt! Der Eine lebte für den Anderen. Du siehst sie überall und in allem was dich umgibt. Überall spürst du ihren Duft. Die Traurigkeit überkommt dich:

- „Es ist Ungerecht. Warum musste gerade einer von uns beiden gehen. Das ist nicht möglich. Ich kann es nicht glauben! Ich kann so nicht weiter leben. Ich brauche dich. Und heute habe ich auch noch mit den Japanern eine Versammlung!“

Appendice VII - Scripts per la somministrazione dei test di validazione T1 e T2 in Praat

Script test di validazione T1

```
"ooTextFile"
"ExperimentMFC 5"
stimuliAreSounds? <yes>
stimulusFileNameHead = "Sounds/"
stimulusFileNameTail = ".mp3"
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1.2 seconds
stimulusMedialSilenceDuration = 0
numberOfDifferentStimuli = 120
"it_ang_a_ms_a1" ""
"it_ang_n_gb_a1" ""
"it_ang_n_gc_c3" ""
...
"it_sad_a_ms_c1" ""
"it_sad_n_gb_c2" ""
"it_sad_n_gc_c1" ""
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 30
randomize = <PermuteBalancedNoDoublets>
startText = "Questo è un esperimento percettivo.
Prima di iniziare verifica i requisiti tecnici
riportati nel readme file.

Ascolterai degli stimoli vocali.
Per ciascuno stimolo ti verrà chiesto di:
1. assegnare un'emozione;
2. indicare quanto sei certo della tua scelta.

Clicca qui per iniziare il test."
runText = "Assegna un'emozione.
Di seguito indica quanto sei certo della tua scelta."
pauseText = "Fai una breve pausa!
Clicca qui per continuare."
endText = "Test 1 completato!

Dopo aver chiuso la finestra,
segui le istruzioni per esportare i risultati."
maximumNumberOfReplays = 2
replayButton = 0.26 0.48 0.11 0.17 "Replay" ""
okButton = 0.50 0.72 0.11 0.17 "Avanti" ""
oopsButton = 0.0 0.0 0.00 0.0 "" ""
responsesAreSounds? <no> "" "" "" 0 0
numberOfDifferentResponses = 7
0.02 0.13 0.60 0.70 "#T#r#i#s#t#e#z#z#a" 20 "" "1"
0.16 0.27 0.60 0.70 "#C#o#l#l#e#r#a" 20 "" "2"
0.30 0.41 0.60 0.70 "#P#a#u#r#a" 20 "" "3"
0.44 0.55 0.60 0.70 "#N#e#u#t#r#o" 20 "" "4"
0.58 0.69 0.60 0.70 "#G#i#o#i#a" 20 "" "5"
0.72 0.83 0.60 0.70 "#S#o#r#o#r#e#s#a" 20 "" "6"
0.86 0.97 0.60 0.70 "#D#i#s#g#u#s#t#o" 20 "" "7"
numberOfGoodnessCategories = 2
0.15 0.37 0.32 0.38 "incerto"
0.63 0.85 0.32 0.38 "certo"
```

Script test di validazione T2

```

"ooTextFile"
"ExperimentMFC 5"
stimuliAreSounds? <yes>
stimulusFileNameHead = "Sounds/"
stimulusFileNameTail = ".mp3"
stimulusCarrierBefore = ""
stimulusCarrierAfter = ""
stimulusInitialSilenceDuration = 1.2 seconds
stimulusMediaSilenceDuration = 0
numberOfDifferentStimuli = 120
"it_ang_a_ms_a1" "Lo stimolo rappresenta l'emozione di #C#O#L#L#E#R#A."
"it_ang_n_gb_a1" "Lo stimolo rappresenta l'emozione di #C#O#L#L#E#R#A."
"it_ang_n_gc_c3" "Lo stimolo rappresenta l'emozione di #C#O#L#L#E#R#A."
...
"it_sad_a_ms_c1" "Lo stimolo rappresenta l'emozione di #T#R#I#S#T#E#Z#Z#A."
"it_sad_n_gb_c2" "Lo stimolo rappresenta l'emozione di #T#R#I#S#T#E#Z#Z#A."
"it_sad_n_gc_c1" "Lo stimolo rappresenta l'emozione di #T#R#I#S#T#E#Z#Z#A."
numberOfReplicationsPerStimulus = 1
breakAfterEvery = 30
randomize = <PermuteBalancedNoDoublets>
startText = "Questo è un esperimento percettivo.
Prima di iniziare verifica i requisiti tecnici
riportati nel readme file.

Ascolterai degli stimoli vocali.
Per ciascuno stimolo ti verrà chiesto di
indicare quanto lo stimolo presentato
rappresenta l'emozione data
(da #0= %p%e%r %n%i%e%n%t%e a #3= %a%p%p%i%e%n%o).

Clicca qui per iniziare il test."
runText = ""
pauseText = "Fai una breve pausa!
Clicca qui per continuare."
endText = "Test 2 completato!

Dopo aver chiuso la finestra,
segui le istruzioni per esportare i risultati."
maximumNumberOfReplays = 2
replayButton = 0.26 0.48 0.11 0.17 "Replay" ""
okButton = 0.50 0.72 0.11 0.17 "Avanti" ""
oopsButton = 0.0 0.0 0.00 0.0 "" ""
responsesAreSounds? <no> "" "" "" "" 0 0
numberOfDifferentResponses = 4
    0.13 0.28 0.50 0.70 "#0 %p%e%r %n%i%e%n%t%e" 20 "" "1"
    0.33 0.48 0.50 0.70 "#1" 20 "" "2"
    0.53 0.68 0.50 0.70 "#2" 20 "" "3"
    0.73 0.88 0.50 0.70 "#3 %a%p%p%i%e%n%o" 20 "" "4"
numberOfGoodnessCategories = 0
    0.15 0.37 0.32 0.38 "0"
    0.39 0.61 0.32 0.38 "1"
    0.63 0.85 0.32 0.38 "2"

```

Appendice VIII - Schermate test percettivo *web-based*

The screenshot shows a web browser window displaying a form for a vocal emotion recognition test. The form is titled "Riconoscimento di emozioni vocali" and contains the following text and fields:

In questo esperimento risponderai dapprima ad alcune domande di carattere anagrafico (data di nascita, sesso ecc.).

Subito dopo inizierà il test percettivo vero e proprio in cui ascolterai delle voci in diverse lingue. Per ognuna di esse ti verrà chiesto di indicare l'emozione espressa in quella voce (fai attenzione a come la frase viene pronunciata e non al suo significato).

Infine ti verranno fatte alcune domande sul test percettivo svolto.

Prima di iniziare il test assicurati di avere un paio di cuffie e di aver alzato il volume delle cuffie.

NB: Il passaggio alla prossima pagina potrebbe richiedere del tempo. Pazienta un pò!

Inserisci il tuo nome e il tuo cognome:

Indica il tuo sesso:

Maschio

Femmina

Quanti anni hai?

Di che nazionalità sei?

At the bottom of the browser window, the text "Perception in licenza all'Università della Calabria" is visible. The browser's status bar shows "Internet | Modalità protetta: attivata" and "100%".

Figura 72: Schermata 1 di 3 del test percettivo somministrato via web con Questionmark™ Perception™

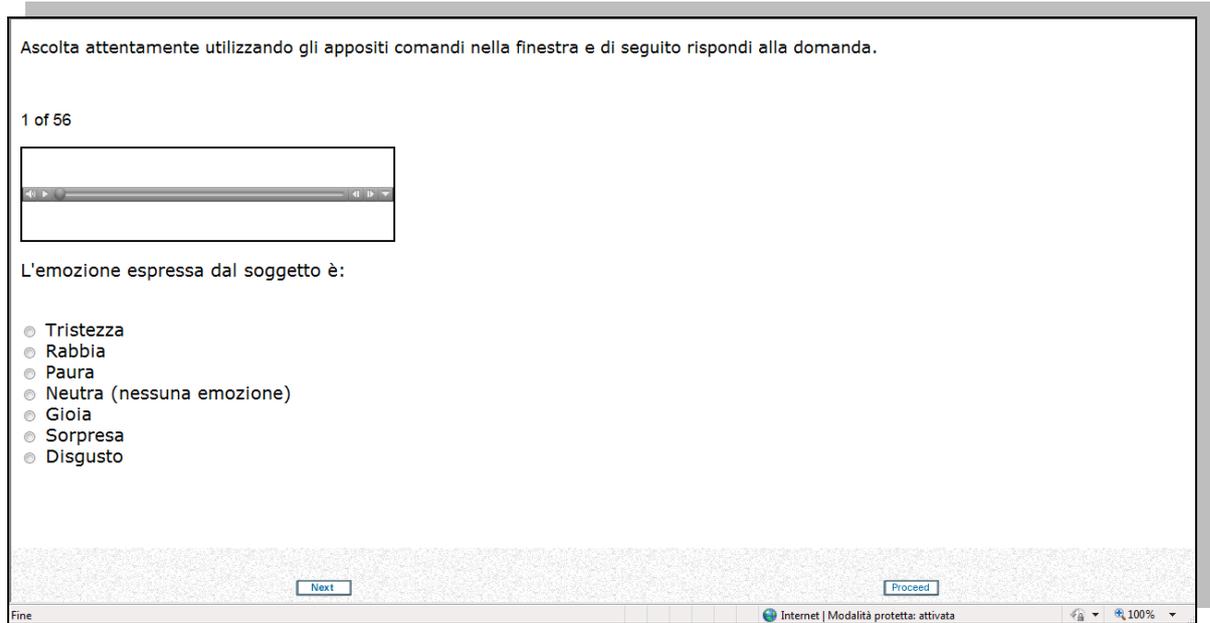


Figura 73: Schermata 2 di 3 del test percettivo somministrato via web con Questionmark™ Perception™

Riconoscimento di emozioni vocali

Rispondi a queste ultime domande.

Per quale lingua hai avuto più difficoltà a riconoscere le emozioni?

Italiano
 Francese
 Inglese
 Tedesco

A prescindere dalla lingua, indica quali sono, secondo te, le due emozioni più difficili da riconoscere.

Tristezza
 Rabbia
 Paura
 Neutra (nessuna emozione)
 Gioia
 Sorpresa
 Disgusto

Indica il tuo grado di competenza nella lingua ITALIANA:

Nessuna
 Apprendista
 Scolastica
 Universitaria
 Avanzata
 Madrelingua

Indica il tuo grado di competenza nella lingua FRANCESE:

Nessuna
 Apprendista
 Scolastica
 Universitaria
 Avanzata
 Madrelingua

Indica il tuo grado di competenza nella lingua INGLESE:

Nessuna
 Apprendista
 Scolastica
 Universitaria
 Avanzata
 Madrelingua

Indica il tuo grado di competenza nella lingua TEDESCA:

Nessuna
 Apprendista
 Scolastica
 Universitaria
 Avanzata
 Madrelingua

Sei in possesso di qualche certificazione (ad es.: TOEFL, Trinity, DALF/DELF, ecc.) per qualcuna delle lingue straniere che qui hai ascoltato?

Yes
 No

Se sì, indica quale/i?

Hai qualche commento o osservazione da fare sull'esperimento a cui hai appena partecipato?

Figura 74: Schermata 3 di 3 del test percettivo somministrato via web con Questionmark™ Perception™

Appendice IX - Comunicazione inoltrata via mail per il reclutamento dei soggetti ascoltatori

Ciao,
mi chiamo Vincenzo Galatà e sono un dottorando in "Psicologia della Programmazione e Intelligenza Artificiale" presso il dipartimento di Linguistica dell'Università della Calabria.

Sto portando avanti una ricerca volta allo studio delle emozioni espresse attraverso la voce a livello cross-linguistico, sia da un punto di vista di produzione che di percezione.

Per indagare su quest'ultimo aspetto, legato alla percezione delle emozioni a livello cross-linguistico, ho costruito un test percettivo che viene somministrato via web attraverso un'interfaccia grafica semplice ed intuitiva. Non sono richiesti particolari requisiti per partecipare all'esperimento se non un po' di tempo libero (da 10 a 20 minuti), un paio di cuffie (anche da iPod), un collegamento a internet... e un pizzico di buona volontà!

Se hai voglia di aiutarmi partecipando a questo interessante esperimento, collegati a questo link ([esperimento](#)) e segui le istruzioni che vengono riportate a schermo. Sulla prima pagina che ti appare scrivi semplicemente il tuo nome e clicca su "Enter". Successivamente clicca sulla voce "Riconoscimento di emozioni vocali" ed inizia l'esperimento vero e proprio seguendo le istruzioni riportate sullo schermo.

PS: Ad un certo punto, dopo le prime quattro domande sulla prima pagina, a seconda del browser dovrebbe essere necessario accettare l'installazione di "controlli ActiveX" (sotto forma di pop-up) o plug-in, fondamentali per la riproduzione e l'ascolto dei files che ti verranno presentati. Se dopo l'installazione di tali "controlli" appare un messaggio di errore, ricomincia il test dall'inizio (dovrai solo rispondere nuovamente a 4 semplici domande come età, sesso ed altro). Il passaggio dalla prima alla seconda pagina potrebbe richiedere del tempo. Pazienta un po' (qualche minuto) e dai il tempo alla pagina di caricare tutti i files.

PS2: è stato risolto qualche problema di visualizzazione... il test funziona ora con diversi browser (Explorer, Safari, Google Chrome), ma NON con Firefox e Opera.

Sei hai dubbi o problemi con il test non esitare a scrivermi, sarò felice di aiutarti.
Inoltre questa mail ai tuoi amici, forse c'è qualcuno a cui piacerebbe partecipare all'esperimento.

Grazie mille!
Vincenzo

Appendice X - Commenti dei rispondenti a fine esperimento

A conclusione dell'esperimento è stata concessa ai rispondenti la possibilità di rilasciare commenti sull'esperimento svolto, alcuni dei quali vengono di seguito riportati in Tabella 44.

Alcuni commenti dei rispondenti

L'identificazione tra i diversi stati emotivi è abbastanza ardua per l'unica lingua non conosciuta (il tedesco). Delle altre tre il limite tra uno stato emotivo e l'altro rimane comunque sottile ma credo di aver riconosciuto le produzioni linguistiche relative agli stati di disgusto - gioia - sorpresa. Di questi ultimi due, inoltre, credo si possa dire che possono essere facilmente accostati.

La scelta è stata difficile per tutte le lingue.

Per disgusto sorpresa e tristezza secondo me è già difficile riconoscere le emozioni proposte per la lingua italiana. Diventa forse ancora più difficile per le altre lingue, soprattutto per quelle relativamente poco familiari (francese, tedesco, in questo caso).

In alcuni casi sarebbe stata utile una risposta tipo "altro", perché l'impressione era di scarsa naturalezza, di stanchezza o di fretta/divertimento.

Credevo fosse più facile e invece... Alcune emozioni erano davvero difficili da interpretare, soprattutto all'inizio del test. Ho avuto maggiori difficoltà con la lingua TEDESCA perché è quella che conosco meno bene. Molto più facili e divertenti le frasi in FRANCESE...

Per quanto cercassi di non tenere conto del significato delle esclamazioni, nelle lingue da me conosciute le emozioni sembravano più facilmente riconoscibili.

Oltre alle indicazioni in griglia è possibile riconoscere anche sfumature che mescolano diversi stati d'animo, per es. gioia e sorpresa insieme, paura e rabbia contemporaneamente ecc.

Alcune frasi mi comunicavano emozioni diverse da quelle elencate, per cui risultava un pò forzato inserirle tra le opzioni possibili.

Qualche problema in più di quanto mi aspettassi a riconoscere le emozioni in parlanti italiani

Mi è sembrato molto strano trovarmi più in difficoltà nell'identificazione di emozioni nella mia madrelingua che non nelle altre lingue presenti nel test.

Manca il riconoscimento di emozioni vocali in lingua spagnola. Aggiungerei anche lingue asiatiche.

Personalmente credo sia poco possibile riuscire a scindere il suono delle parole dal loro senso. Tuttavia penso che, eccetto le emozioni quali la tristezza, la rabbia e la gioia, (le più semplici da riprodurre), le altre elencate nel test siano molto relative. Ovvero, è relativo il genere di musicalità che ciascuno conferisce ad un'emozione quale il disgusto, la sorpresa o il suono neutro, per esempio. Inoltre un'emozione costruita (in questo caso dal timbro di una voce) appare immediatamente artefatta, cioè non autentica, di conseguenza, probabilmente, poco valida. Ma certo questo esperimento avrà il suo senso compiuto. Mi ha fatto piacere, dunque, potervi partecipare.

E' stato interessante ed anche molto divertente, ma non nego il fatto di avere avuto difficoltà molto spesso nel riconoscere le emozioni.

Mi mancava qualche emozione nel catalogo, tipo disperazione, fastidio, incredulità, indifferenza

Tabella 44: Commenti dei rispondenti all'esperimento di riconoscimento di emozioni vocali svolto

Appendice XI - Lo script per l'estrazione dei parametri in Praat

```

### SCRIPT BY VINCENZO GALATA'
### vgalata@libero.it

form with this script you can analyze labeled files (F0 etc.)
  comment Directory del file C:\...\
  text sound_directory \
  comment Directory delle etichette C:\...\
  text textGrid_directory \
  sentence Sound_file_extension .wav
  sentence TextGrid_file_extension .TextGrid
  comment Quale tier del TextGrid vuoi analizzare?
  integer Tier 1
comment inserisci il genere del parlatore
  sentence genere m
comment inserire il nome del file
  text nome_file
endform
directory$ = sound_directory$

  header_row$ = "Filename" + tab$ + "Language" + tab$ + "Emotion" + tab$ + "Type" + tab$
+ "Subject" + tab$ + "Segment" + tab$ + "Durata (ms.)" + tab$ + "F0_min" + tab$ + "F0_max"
+ tab$ + "F0_mean" + tab$ + "F0_stdev" + tab$ + "F0_range" + tab$ + "F0_median" + tab$ +
"F0_q25" + tab$ + "F0_q75" + tab$ + "Int_min" + tab$ + "Int_max" + tab$ + "Int_mean" + tab$ +
+ "Int_stdev" + tab$ + "Int_range" + tab$ + "Int_median" + tab$ + "Int_q25" + tab$ +
"Int_q75" + newline$
  header_row$ > 'directory$\'nome_file$.xls

  Create Strings as file list... list 'directory$\'*.wav
  number_files = Get number of strings

pause Check Pitch and Intensity settings before going on!

for j from 1 to number_files

  select strings list
  current_token$ = Get string... 'j'
  Read from file... 'directory$\'current_token$'

  object_name$ = selected$ ("Sound")

  select Sound 'object_name$'
  if genere$ = "m"
    To Pitch... 0.01 60 300
  else
    To Pitch... 0.01 75 600
  endif
  select Sound 'object_name$'
  To Intensity... 60 0 yes

select sound 'object_name$'
language$= left$(object_name$, 2)

  if language$ = "it"
    language$= "italian"
  elseif language$ = "fr"
    language$= "french"
  elseif language$ = "en"
    language$= "english"
  elseif language$ = "de"
    language$= "german"
  endif

emotion$= mid$(object_name$, 4, 3)

  if emotion$ = "ang"
    emotion$= "anger"
  elseif emotion$ = "dis"
    emotion$= "disgust"
  elseif emotion$ = "fea"
    emotion$= "fear"
  elseif emotion$ = "neu"
    emotion$= "neutral"
  elseif emotion$ = "sad"

```

```

        emotion$= "sadness"
    elseif emotion$ = "sur"
        emotion$= "surprise"
    elseif emotion$ = "joy"
        emotion$= "joy"
    endif

    actor$= mid$(object_name$, 8, 1)
    if actor$ = "a"
        actor$= "actor"
    elseif actor$ = "n"
        actor$= "naif"
    endif

    subject$= mid$(object_name$, 10, 2)

    Read from file... 'directory$('object_name$'.TextGrid

    select TextGrid 'object_name$'
    number_of_intervals = Get number of intervals... tier
    for b from 1 to number_of_intervals
        select TextGrid 'object_name$'
        text$ = Get label of interval... tier 'b'
        if text$ <> ""
            t1 = Get starting point... tier 'b'
            t2 = Get end point... tier 'b'
            select Pitch 'object_name$'
            f0_max = Get maximum... t1 t2 Hertz None
            f0_min = Get minimum... t1 t2 Hertz None
            f0_mean = Get mean... t1 t2 Hertz
            f0_median = Get quantile... t1 t2 0.5 Hertz
            f0_range = f0_max - f0_min
            f0_q25 = Get quantile... t1 t2 0.25 Hertz
            f0_q75 = Get quantile... t2 t2 0.75 Hertz
            f0_stdev = Get standard deviation... t1 t2 Hertz
            duration = (t2 - t1)
            select Intensity 'object_name$'
            int_min = Get minimum... t1 t2 Parabolic
            int_max = Get maximum... t1 t2 Parabolic
            int_mean = Get mean... t1 t2 dB
            int_stdev = Get standard deviation... t1 t2
            int_range = int_max - int_min
            int_median = Get quantile... t1 t2 0.5
            int_q25 = Get quantile... t1 t2 0.25
            int_q75 = Get quantile... t1 t2 0.75

                fileappend "'directory$('nome_file$'.xls"
            'object_name$('tab$('language$('tab$('emotion$('tab$('actor$('tab$('subject$('tab$('text$('
            tab$('duration:3' tab$('f0_min:2' tab$('f0_max:2' tab$('f0_mean:2' tab$('f0_stdev:2' tab$('
            f0_range:2' tab$('f0_median:2' tab$('f0_q25:2' tab$('f0_q75:2' tab$('int_min:2' tab$('int_m
            ax:2' tab$('int_mean:2' tab$('int_stdev:2' tab$('int_range:2' tab$('int_median:2' tab$('int
            _q25:2' tab$('int_q75:2' 'newline$'
            endif
        endfor

        select all
        minus Strings list
        Remove
    endfor

    select all
    Remove
    clearinfo
    print          All files processed!!! Your results are in "'directory$('nome_file$'.xls"

    ## adapted by @speedy@

```

Appendice XII - Parametri acustici estrapolati per le produzioni degli attori

Language	Emotion	Type	Dur ms.	F0_min	F0_max	F0_mean	F0_stdev	F0_range	F0_median	F0_q25	F0_q75	Int_min	Int_max	Int_mean	Int_stdev	Int_range	Int_median	Int_q25	Int_q75	Syll	SR	Pause
english	anger	actor	0,872	77,42	217,86	139,57	33,67	140,45	144,42	120,73	158,06	21,72	65,47	49,39	10,96	43,75	51,74	42,17	58,23	4,5	5,03	0,561
english	disgust	actor	1,140	83,26	192,46	140,24	32,89	109,20	133,98	111,11	157,35	13,30	58,47	41,08	14,08	45,18	46,12	31,20	51,48	4,5	3,85	0,000
english	fear	actor	0,877	147,71	262,99	210,92	29,80	115,27	213,79	194,07	233,46	19,69	60,46	46,85	11,56	40,77	49,91	39,78	56,50	4,5	5,00	0,569
english	joy	actor	0,980	89,93	298,80	207,52	59,99	208,87	213,95	165,38	263,98	14,94	62,88	47,75	13,29	47,95	51,33	40,83	58,60	4,5	4,47	0,394
english	sadness	actor	1,081	84,21	171,49	124,39	24,31	87,29	123,78	104,49	141,64	14,17	55,20	36,52	13,73	41,03	40,02	23,43	48,45	4,5	4,12	0,661
english	surprise	actor	0,954	93,63	294,80	204,59	66,05	201,18	215,21	136,50	266,20	15,55	61,50	44,89	12,82	45,95	48,25	38,38	55,25	4,5	4,62	0,469
english	neutral	actor	0,856	72,44	139,49	114,96	21,66	67,05	121,67	94,70	123,96	15,58	55,14	40,07	10,83	39,56	41,50	33,24	48,95	4,5	5,11	0,798
italian	anger	actor	1,134	67,65	239,73	161,55	54,13	172,08	176,76	103,15	209,29	32,64	67,79	58,38	7,61	35,15	59,92	53,95	64,21	6,5	5,72	0,228
italian	disgust	actor	1,448	67,00	131,19	97,84	18,71	64,19	100,86	78,34	113,47	22,36	55,94	46,40	7,09	33,58	48,01	43,61	51,10	6,5	4,50	0,407
italian	fear	actor	0,976	70,85	163,28	112,91	29,03	92,44	112,10	83,32	134,65	29,89	59,47	48,96	7,08	29,58	50,68	43,79	54,02	6,5	6,67	0,511
italian	joy	actor	1,013	103,56	254,05	175,59	43,37	150,49	178,46	131,62	203,42	26,88	62,11	53,37	7,54	35,22	54,55	51,23	58,21	6,5	6,55	0,000
italian	sadness	actor	0,985	67,79	125,08	88,86	12,30	57,29	88,38	82,56	92,44	20,76	48,10	38,24	7,03	27,35	40,52	33,36	43,71	6,5	6,63	1,099
italian	surprise	actor	1,009	63,12	138,52	91,16	22,57	75,41	81,77	73,70	93,51	28,93	53,63	43,80	6,58	24,70	44,44	38,34	49,63	6,5	6,45	0,237
italian	neutral	actor	1,231	69,44	115,63	85,76	13,55	46,19	81,57	74,60	95,16	18,17	50,79	40,44	7,08	32,63	42,07	38,57	44,79	6,5	5,28	0,437
german	anger	actor	1,249	106,07	288,64	192,25	48,58	182,57	191,60	154,70	233,71	30,37	67,49	54,78	7,24	37,13	56,22	50,46	59,41	5,5	4,41	0,000
german	disgust	actor	1,519	76,55	191,65	135,91	25,92	115,09	138,59	122,17	153,96	26,44	63,48	51,80	8,04	37,04	53,38	46,55	58,42	5,5	3,62	0,000
german	fear	actor	1,067	131,26	285,48	208,80	37,22	154,22	207,86	186,72	229,30	37,70	66,83	57,00	6,34	29,13	57,79	54,37	61,24	5,5	5,16	0,178
german	joy	actor	1,110	123,24	294,26	207,60	45,21	171,01	202,01	181,09	233,87	38,58	67,12	57,41	6,68	28,56	58,70	53,64	62,56	5,5	4,96	0,000
german	sadness	actor	1,277	93,47	165,60	117,54	13,98	72,13	116,08	110,57	123,14	22,18	55,71	45,71	7,83	33,53	47,91	40,55	51,64	5,5	4,31	0,992
german	surprise	actor	1,317	85,54	242,31	150,37	42,11	156,77	135,78	124,81	199,18	28,47	57,80	48,98	6,00	29,33	49,82	46,29	53,64	5,5	4,17	0,000
german	neutral	actor	1,256	98,52	141,08	118,17	11,14	42,55	117,76	108,71	129,36	24,72	58,24	48,48	7,36	33,53	49,64	46,23	54,10	5,5	4,37	0,271
french	anger	actor	0,918	129,62	254,28	197,58	36,24	124,67	198,84	171,86	231,17	45,82	69,17	61,46	5,73	23,35	63,30	57,24	65,59	4,0	4,39	0,000
french	disgust	actor	1,150	71,09	201,31	106,61	23,40	130,22	105,53	94,14	112,67	25,08	53,57	44,72	8,25	28,50	47,74	40,66	50,97	4,0	3,43	0,313
french	fear	actor	0,689	165,15	248,53	207,54	26,35	83,37	212,75	183,88	235,29	35,05	61,19	54,10	6,10	26,14	56,04	51,52	57,85	4,0	5,99	0,334
french	joy	actor	1,105	140,22	251,10	203,15	30,21	110,89	211,50	185,87	225,51	45,60	68,73	61,88	5,33	23,13	62,63	58,49	66,53	4,0	3,61	0,000
french	sadness	actor	0,854	90,92	121,57	99,52	5,66	30,64	99,81	96,38	101,92	27,06	47,65	42,37	4,66	20,60	43,15	40,49	46,27	4,0	4,62	0,226
french	surprise	actor	0,776	99,69	175,20	136,80	22,19	75,52	131,14	117,44	155,12	33,13	54,27	48,01	4,53	21,14	49,32	46,21	50,96	4,0	5,11	0,000
french	neutral	actor	0,873	82,63	125,07	98,07	11,11	42,44	95,94	89,21	104,83	27,65	51,16	45,21	6,10	23,52	47,32	44,08	49,22	4,0	4,71	0,927

Tabella 45: Misure medie estrapolate per le produzioni utilizzate nel test percettivo, limitatamente agli encoder appartenenti alla tipologia attore

Appendice XIII - Indici di correlazione tra i parametri acustici estrapolati

		Correlazioni							
			F0_max	F0_mean	F0_stdev	F0_range	F0_median	F0_q25	F0_q75
Italiano	F0_min	Correlazione di Pearson	,676	,714	,437	,507	,648	,887	,612
		Sig. (2-code)	,096	,071	,327	,246	,116	,008	,144
	F0_max	Correlazione di Pearson		,995	,953	,978	,983	,921	,980
		Sig. (2-code)		,000	,001	,000	,000	,003	,000
	F0_mean	Correlazione di Pearson			,934	,961	,992	,940	,986
		Sig. (2-code)			,002	,001	,000	,002	,000
	F0_stdev	Correlazione di Pearson				,991	,942	,762	,961
		Sig. (2-code)				,000	,002	,047	,001
	F0_range	Correlazione di Pearson					,966	,825	,973
		Sig. (2-code)					,000	,022	,000
	F0_median	Correlazione di Pearson						,911	,995
		Sig. (2-code)						,004	,000
	F0_q25	Correlazione di Pearson							,880
		Sig. (2-code)							,009
Francese	F0_min	Correlazione di Pearson	,759	,942	,604	,250	,946	,952	,932
		Sig. (2-code)	,048	,001	,151	,589	,001	,001	,002
	F0_max	Correlazione di Pearson		,910	,947	,820	,908	,889	,909
		Sig. (2-code)		,004	,001	,024	,005	,007	,005
	F0_mean	Correlazione di Pearson			,826	,525	,998	,995	,998
		Sig. (2-code)			,022	,227	,000	,000	,000
	F0_stdev	Correlazione di Pearson				,878	,810	,784	,843
		Sig. (2-code)				,009	,027	,037	,017
	F0_range	Correlazione di Pearson					,518	,486	,532
		Sig. (2-code)					,234	,269	,219
	F0_median	Correlazione di Pearson						,999	,992
		Sig. (2-code)						,000	,000
	F0_q25	Correlazione di Pearson							,987
		Sig. (2-code)							,000
Inglese	F0_min	Correlazione di Pearson	,484	,685	,029	,086	,646	,862	,523
		Sig. (2-code)	,271	,090	,952	,854	,117	,013	,228
	F0_max	Correlazione di Pearson		,953	,860	,914	,955	,793	,976
		Sig. (2-code)		,001	,013	,004	,001	,034	,000
	F0_mean	Correlazione di Pearson			,735	,767	,994	,901	,976
		Sig. (2-code)			,060	,044	,000	,006	,000
	F0_stdev	Correlazione di Pearson				,966	,760	,397	,859
		Sig. (2-code)				,000	,048	,378	,013
	F0_range	Correlazione di Pearson					,786	,502	,868
		Sig. (2-code)					,036	,251	,011
	F0_median	Correlazione di Pearson						,877	,981
		Sig. (2-code)						,010	,000
	F0_q25	Correlazione di Pearson							,796
		Sig. (2-code)							,032
Tedesco	F0_min	Correlazione di Pearson	,646	,802	,400	,403	,814	,867	,649
		Sig. (2-code)	,117	,030	,374	,370	,026	,012	,115
	F0_max	Correlazione di Pearson		,962	,948	,959	,927	,898	,988
		Sig. (2-code)		,001	,001	,001	,003	,006	,000
	F0_mean	Correlazione di Pearson			,841	,855	,990	,979	,959
		Sig. (2-code)			,018	,014	,000	,000	,001
	F0_stdev	Correlazione di Pearson				,988	,786	,727	,951
		Sig. (2-code)				,000	,036	,064	,001
	F0_range	Correlazione di Pearson					,810	,756	,944
		Sig. (2-code)					,027	,050	,001
	F0_median	Correlazione di Pearson						,980	,923
		Sig. (2-code)						,000	,003
	F0_q25	Correlazione di Pearson							,887
		Sig. (2-code)							,008

Tabella 46: Indici di correlazione di Pearson tra gli indici estrapolati nel dominio della frequenza (per ciascun parametro n = 7)

			Correlazioni						
			Int_max	Int_mean	Int_stdev	Int_range	Int_median	Int_q25	Int_q75
Italiano	Int_min	Correlazione di Pearson	,803	,788	,253	,037	,770	,654	,822
		Sig. (2-code)	,030	,035	,584	,938	,043	,111	,023
	Int_max	Correlazione di Pearson		,997**	,724	,626	,996**	,970**	,994**
		Sig. (2-code)		,000	,066	,133	,000	,000	,000
	Int_mean	Correlazione di Pearson			,739	,640	,998**	,979**	,997**
		Sig. (2-code)			,058	,121	,000	,000	,000
	Int_stdev	Correlazione di Pearson				,882**	,771	,806	,713
		Sig. (2-code)				,009	,042	,029	,072
	Int_range	Correlazione di Pearson					,661	,770	,591
		Sig. (2-code)					,106	,043	,163
	Int_median	Correlazione di Pearson						,977**	,994**
		Sig. (2-code)						,000	,000
	Int_q25	Correlazione di Pearson							,962**
		Sig. (2-code)							,001
Francese	Int_min	Correlazione di Pearson	,948**	,977**	-,323	-,183	,955**	,980**	,951**
		Sig. (2-code)	,001	,000	,479	,694	,001	,000	,001
	Int_max	Correlazione di Pearson		,991**	-,020	,138	,998**	,968**	,996**
		Sig. (2-code)		,000	,965	,768	,000	,000	,000
	Int_mean	Correlazione di Pearson			-,140	,023	,995**	,991**	,992**
		Sig. (2-code)			,764	,962	,000	,000	,000
	Int_stdev	Correlazione di Pearson				,946**	-,052	-,232	-,032
		Sig. (2-code)				,001	,912	,617	,946
	Int_range	Correlazione di Pearson					,108	-,060	,117
		Sig. (2-code)					,817	,899	,802
	Int_median	Correlazione di Pearson						,979**	,994**
		Sig. (2-code)						,000	,000
	Int_q25	Correlazione di Pearson							,970**
		Sig. (2-code)							,000
Inglese	Int_min	Correlazione di Pearson	,623	,701	-,765	-,220	,580	,656	,598
		Sig. (2-code)	,135	,080	,045	,635	,172	,109	,156
	Int_max	Correlazione di Pearson		,941**	-,204	,626	,953**	,862	,953**
		Sig. (2-code)		,002	,660	,133	,001	,013	,001
	Int_mean	Correlazione di Pearson			-,400	,475	,971**	,970**	,974**
		Sig. (2-code)			,375	,281	,000	,000	,000
	Int_stdev	Correlazione di Pearson				,508	-,188	-,502	-,211
		Sig. (2-code)				,244	,687	,251	,650
	Int_range	Correlazione di Pearson					,609	,420	,592
		Sig. (2-code)					,146	,349	,161
	Int_median	Correlazione di Pearson						,913**	,981**
		Sig. (2-code)						,004	,000
	Int_q25	Correlazione di Pearson							,911**
		Sig. (2-code)							,004
Tedesco	Int_min	Correlazione di Pearson	,811	,932**	-,635	-,597	,911**	,952**	,884**
		Sig. (2-code)	,027	,002	,125	,157	,004	,001	,008
	Int_max	Correlazione di Pearson		,966**	-,197	-,015	,976**	,892**	,969**
		Sig. (2-code)		,000	,672	,975	,000	,007	,000
	Int_mean	Correlazione di Pearson			-,390	-,268	,995**	,968**	,985**
		Sig. (2-code)			,387	,561	,000	,000	,000
	Int_stdev	Correlazione di Pearson				,816	-,316	-,550	-,252
		Sig. (2-code)				,025	,490	,201	,586
	Int_range	Correlazione di Pearson					-,219	-,404	-,181
		Sig. (2-code)					,637	,369	,697
	Int_median	Correlazione di Pearson						,941**	,990**
		Sig. (2-code)						,002	,000
	Int_q25	Correlazione di Pearson							,924**
		Sig. (2-code)							,003

Tabella 47: Indici di correlazione di *Pearson* tra gli indici estrapolati nel dominio dell'intensità (per ciascun parametro n = 7)