Languaging Diversity Volume 2:

Variationist Approaches and Identities

Edited by

Siria Guzzo and David Britain

Cambridge
Scholars
Publishing

TABLE OF CONTENTS

EXPRESSING AND LABELLING EMOTIONS IN A SECOND LANGUAGE: ACOUSTIC AND PERCEPTUAL ANALYSES OF EMOTIONAL SPEECH IN L1-L2 ITALIAN

ELISA PELLEGRINO AND MARTA MAFFIA
UNIVERSITY OF NAPLES L'ORIENTALE

1. Introduction

The expression of emotions and their recognition represent a crucial aspect in everyday communication. Human beings can rely on different channels to express their emotions, including voice, spoken words, non verbal vocal expressions and body language. Changes in breathing, posture, facial expressions and acoustic signals are indeed indicative of the speaker's emotional state. However, among these cues, the human voice is one of the most powerful and suitable tools for the identification of the speaker. From speech, listeners can gather information about the interlocutor's age, sex, body size, and geographical origin (Hartmann and Danahuer 1976; Mullennix 1995; Linville 1996; Karpf 2006; Pettorino, Pellegrino and Maffia 2014), as well as information about his/her emotional state. According to Belin, Fecteau and Be´dard (2004), the human voice represents an 'auditory face', that allows the identification of an individual's emotional and motivational state. Several acoustic parameters, such as level, range and contours of fundamental frequency (F0), intensity, voice quality, duration and spectral characteristics of the utterance are indeed influenced by (Poggi and Magno Caldognetto 2004).

Many intra- and cross-linguistic studies have been conducted on the identification of the acoustic correlates of basic emotions: joy, fear, anger, sadness, disgust, and surprise (Ekman 1992) (see Johnstone and Scherer 2000; Scherer 2003 for a review). Nonetheless, despite a great effort to categorize emotions, there is a considerable divergence of opinions about what method would be the most suitable for collecting, labeling and

analyzing expressive speech. This lack of agreement among researchers represents a huge obstacle to the scientific study of emotions and, at the same time, makes the identification of shared emotional speech patterns unattainable.

Owing to these methodological and analytical problems, the first issue to be addressed when dealing with the collection of expressive spoken corpora is the choice of the style of speech to analyze. According to Scherer's taxonomy, it is possible to choose between three kinds of performance: play-acted expressions, authentic emotional speech, induced affective speech (Scherer 2003). Each of these modalities of conveying emotions has both strengths and weaknesses.

For example, those who concentrate on play-acted expressions assume that they are closely related to authentic expressions and have the advantage that the verbal and phonetic contents of the performance are easily controlled. The main points of criticism of this modality are that actors cannot really imitate the physiological reactions that accompany emotions and may produce stereotypic forms of emotions influenced by display rules (see Jürgens, Kurt and Fischer 2011 for a review).

The use of spontaneous emotional speech, on the other hand, guarantees data authenticity but, at the same time, this procedure does not allow the collection of comparable data or control of the large numbers of variables affecting spoken communication.

Finally, the so-called Mood Induction Procedures enable the researcher to have a certain degree of control over the stimulus and the context in which emotions are produced (Gerrard-Hesse, Spies and Hesse 1994).

Other methodological issues to be addressed when dealing with emotional speech are: the choice of speakers (actors or non professional actors), the kind of stimulus (for example pictures, videos, dyadic interactions) and the linguistic materials (nonsense words, syllables, interjections, utterances) (see Coan and Allen, 2007 for a review of the most common methods for emotion elicitation).

In addition to these methodological difficulties, another main challenge in emotion detection, particularly in the case of authentic expressive speech, is the categorization and annotation phase (Vidrascu and Devillers 2005). Indeed, the research findings can vary according to whether emotions are conceptualized either as discrete categories (Anger, Fear, Disgust…) or along a continuum of three abstract dimensions: Arousal (Activation), Valence (Pleasure), and Control (Power) (Osgod, May and Miron 1975). As regards annotation, there are at least three different methods of labelling emotions. One first procedure is the annotation by human experts trained to deduce labelling paradigms from theoretical

hypotheses on the nature of emotions. The limitations of this technique, though presumed to be the most objective, have been underlined by Aubergé, Audibert and Rilliard (2006). The most serious objection that the authors raised against this procedure is that the experts may annotate emotions according to their labelling competence and previous experiences. Conversely, they proposed an alternative method, auto-annotation, in which the speakers themselves are asked to judge their own emotional state. According to Aubergé, Audibert and Rilliard (2006) this procedure presents two main advantages. On the one hand it prevents the expert from "taking part" in the situation, on the other, it preserves the correspondence "between the subjectivity of labeling and the subjectivity of the subject him/her-self" (Aubergé, Audibert and Rilliard, 2006:44-45). A third labelling option is the administration of perceptual tests to naïve listeners. However, this procedure is very subjective and may lead to confusing results. Therefore, in order to verify the validity of emotional labels, it would be better to cumulate the findings from different annotation methods (Truong et al. 2009).

Despite the great variability in the ways of collecting and annotating expressive speech, there seems to be a demonstrated correlation between the activation dimension and the most frequently measured acoustic parameters (Schroeder et al. 2001; Schroeder 2003). High activation emotions (fear, joy, surprise and anger) are generally characterized by shorter pauses, wider tonal range, higher values of F0 and intensity, and faster speech rate. Low activation emotions (sadness and disgust), by contrast, are transmitted by longer pauses, narrower tonal range, lower values of F0 and intensity, and slower speech rate. Moreover, cross-linguistic studies on emotional speech encoding and decoding have emphasized the role of prosodic features in the identification of different emotional categories and have indicated that specific emotional states are transmitted by universal prosodic patterns (Yang and Campbell 2001; Pfitzinger et al. 2011).

2. Expressing emotion in the L2

The bulk of studies on the acoustic correlates of emotions have focused on the inter- and intra-language variability in the expression of affection. However, the acoustic correlates of L2 emotional speech have not yet been extensively analyzed. Research on expressive interlanguage has focused on the use of emotion vocabulary by learners of L2 English with different mother tongues (Dawaele and Pavlenko 2002; Kaneko 2003). The objective of these studies was to investigate the emotional force of

swearwords, taboo words or love words when produced or heard in a second language (Dawaele 2003, 2008).

The expression of emotions in a second language was also the object of psycho-physiological investigations conducted by Harris, Aycicegi and Gleason (2003). The study focused on the impact of emotional expressions in L1 and L2 on bilinguals. They monitored the skin conductance of Turkish–English bilinguals via fingertip electrodes while they were rating a variety of stimuli in Turkish (L1) and English (L2) for pleasantness. Participants demonstrated greater arousal to taboo words and childhood reprimands when addressed in their L1 than in their L2.

Further studies on bilingualism demonstrated that the difference between L1 and L2 emotional forces is not so noticeable in early bilinguals (Harris 2004; Harris, Aycicegi and Gleason 2006). Harris' hypothesis, therefore, is that L1 is the language of emotional expressiveness, while L2 is that of emotional distance.

3. The study

Although expressive interlanguage has been examined from different points of view, as we have already mentioned above, the acoustic correlates of L2 emotional speech have not been extensively investigated. The objective of the present study is firstly to identify whether there are some acoustic differences between L1 and L2 expressive speech. In doing so, we will experiment the Card Task, which is a new procedure to collect emotional speech. If compared to the three methods for emotion elicitation mentioned above, this technique has the advantage of controlling the stimuli and the context of interaction. Moreover, it preserves the spontaneity of emotional speech reactions. A further objective is to test the perceptual effect of the acoustic features of expressive speech in L2 Italian both on native and non-native listeners.

3.1. The Card Task

Similarly, the Map Task technique (Anderson et al, 1991), Card Task consists of a non-linguistic activity where participants work in pairs. They sit at the same table but they cannot see each other because of the presence of a dividing panel. One of the participants, the Giver, receives five cards in a given order. Each card has a different drawing on it (fig. 1). The Follower, instead, receives a deck of 25 cards, including four of the five cards owned by the Giver.

Fig. 1: Giver's cards.

The Giver describes the drawing on the first card as accurately as possible. The Follower has two minutes to find the corresponding card in the deck. The same procedure is repeated for the other four cards. The Follower's task is rather difficult because the deck contains very similar cards differing only very slightly from each other (fig. 2).
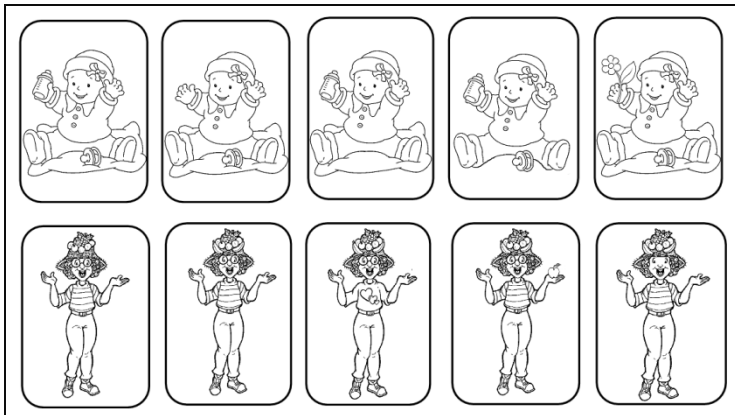
Fig. 2: Examples of some cards in the Follower's deck.

The experimenter is in the same room. She has the function of ensuring that the task is performed as planned and also of eliciting the target linguistic reactions from the players. It is important to underline, however, that the emotions are not elicited by the drawings on the cards but by a series of unexpected events occurring during the game, such as:

1. at the very beginning, while the Follower is looking for the first card in the deck, the stopwatch rings after only 20 seconds instead of after the expected two minutes;

2. the experimenter pretends to accidentally find a big beetle, (obviously fake), that was planted in the room before the game started;

3. the experimenter tries to leave the room but the emergency exit door appears to be locked;

4. the fifth card is not in the deck.

These events were intended to elicit five different emotional reactions from the participants, such as anger, fear, disgust, anxiety and surprise. The unexpected events are resolved in a very short time (only a matter of a few seconds) in order to avoid exaggerated emotional reactions to the situation.

In this study, the Card Task game was organized in two sessions: the first involving only Italian speakers, the second only Chinese participants. The interactions were performed in the silent chamber at "L'Orientale", University of Naples and were videotaped.

3.2. Participants

In the first phase of the study, 10 Italians and 10 Chinese learners of L2 Italian (C1 Level – CEFR) were involved. They were all female, university students, aged between 18 and 23. Before the game, they all declared that they were in good physical condition.

We involved Chinese learners in the experiment because, as shown in the relevant literature, Italian and Chinese speakers differ in the way they express the same emotions, because of their different cultural standards. According to Anolli et al. (2008), Chinese speakers, being more inhibited than Italians, rely mainly on speech rate to vocally communicate emotions. Conversely, Italian speakers, who have a more expansive behavior, accord greater emphasis to pitch and intensity.

As for the level of linguistic competence, we decided to recruit advanced learners (C1) in order to collect a large corpus of emotional speech in L2 Italian and to prevent performance anxiety, caused by inadequate competence, from affecting the study results.

To encourage competitive participation in the game, the task was presented as an official prize competition organized by the "Alumni Association" of "L'Orientale", University of Naples.

At the end of the task, none of the participants expressed any distress. Participants were also given an explanation as to the real purpose of the recordings and then awarded a prize.

4. Methods

4.1. Acoustic analysis

In order to identify the emotional responses to the stimuli presented by the experimenter during the task, we extracted the utterances occurring in the time interval between the end of each stimulus and the resolution of the unexpected events. A total of 132 utterances of expressive speech were collected and then, by means of Praat, the whole corpus was segmented and annotated in two tiers: syllables and speech runs.

For each speech portion, we measured the duration and number of syllables, the length of burst phenomena, silent and filled pauses, and the lowest and highest F0 values. We consider bursts as "very brief, discrete, non-verbal expressions of affection in both face and voice triggered by clearly identifiable events" (Scherer 1994). On the basis of these measurements, we calculated the following indices: articulation rate (AR) (syll./s), speech time composition (percentage of silence, disfluence, syllable and burst), and tonal range (st). Additionally, in order to highlight the F0 variations connected to the emotions considered, for every utterance we related the F0 min and max values to the lowest F0 value (the F0 floor) reached by the speaker in the whole corpus.

4.2. Auto-annotation

The 132 emotional utterances extracted were firstly the object of auto-annotation. During the session, the Italian and Chinese participants in the Card Task were instructed to watch their video recordings and to annotate them with one of the six labels, the five expected emotions, plus a generic option "other". In the case of auto-annotation, the video was supposed to help the players contextualize the utterances and recall what they were feeling during the task.

In order to prevent Chinese participants from misunderstanding the relationship between the labels of emotions and referents of these labels, the emotion categories were translated to their L1 by a native Chinese speaker, who specialized in Italian language and linguistics.

The acoustic data presented in the following paragraph are organized on the basis of the auto-annotation labels.

5. Results

5.1. Spectro-acoustic data

Figure 3 illustrates the mean values of tonal range and F0 height (register) for each emotion in L1 and L2 Italian.
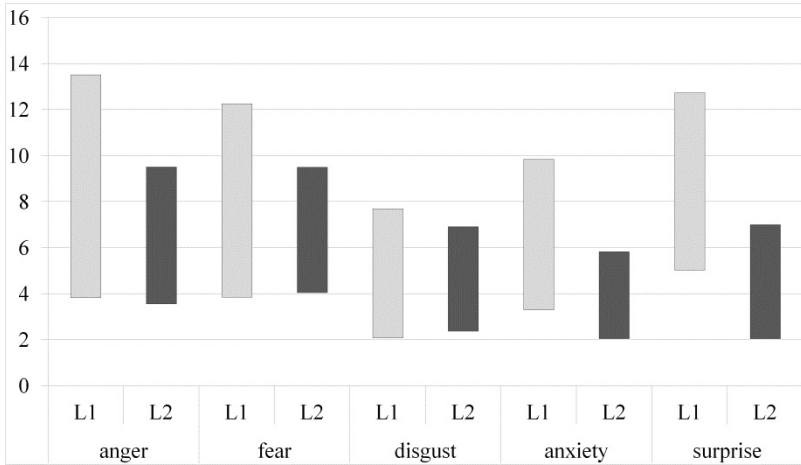

Fig. 3: Tonal Range and F0 height in L1 and L2 emotions.

As the chart shows, the values regarding L1 Italian confirm the patterns for the high and low activation emotions, mentioned in the relevant literature. Anger, fear and surprise correlate with the highest F0 values and the widest tonal range. Anxiety, another high activation emotion, presents pitch height and tonal range values that are slightly lower than surprise, but higher than disgust, a low activation emotion (Jones, Anagnostou and Verhoeven 2011).

Shifting the attention to the acoustic correlates of the emotions expressed by the Chinese subjects, it is possible to observe that the F0 values match those attained by native speakers. As a matter of fact, F0 height and tonal range are quite steady in the whole corpus. The only exception is represented by anger and fear that are expressed with slightly higher values. This data seems to suggest that Chinese learners do not vary their pitch contour to distinguish different emotional states unlike native Italian speakers. After all, smaller pitch excursions are not peculiar only to L2 emotional speech, but they also represent one of the main acoustic

correlates of Chinese accented Italian (De Meo and Pettorino, 2011a, 2011b).

Another parameter under study was articulation rate. Table 1 shows mean values for each emotional state in L1 and L2 Italian.

	Anger	Fear	Disgust	Anxiety	Surprise
L1 Italian	6.2	5.6	5.5	6.1	6.4
L2 Italian	5.2	5.6	4.3	4.8	4.8

Table 1: Articulation rate (syll/s) in L1 and L2 Italian.

Before considering AR variations in expressive speech, it is worth underlining that AR is quite a stable parameter the variations of which are limited by the anatomical and physiological constraints of the phonatory organs. With the exception of fear where L1 and L2 speakers reach the same values, in the other cases, non-native utterances are produced with a slowing down of 1 or 1.5 syll/s, thus confirming the data available in the literature on foreign accented Italian (Pellegrino 2012). The lower values of AR in L2 are essentially due to the learners' greater accuracy in uttering single vowels and consonants. In order to reach the articulatory targets, lengthening of syllable duration is needed, with a consequent slowing down of speech.

Nevertheless, variations in AR do not seem to correlate with the five target emotions either in L1 or in L2. The only exception is represented by disgust in L2 Italian, whose values are particularly low (4.3 syll/s) with respect to the whole corpus.

Figure 4 shows the composition of the utterance for each emotion in the two groups of participants.
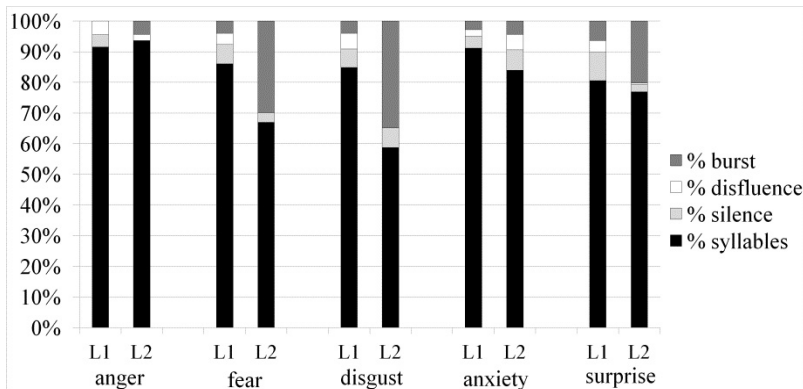

Fig. 4: Speech time composition in L1 and L2 Italian.

In L1 Italian, the percentage of syllables ranges from 85% to 95%, while in the second language it changes in the range of 60% and 93%. In the emotions that mostly differentiate L1 and L2 (fear, disgust and surprise), the bursts increase at the expense of syllable percentage. This means that when non-native speakers are not able to verbalize their emotions, bursts substitute the textual component. However, non-verbal expressions are not equally distributed between the five emotions. For example, in anger they are practically absent. On the contrary, in disgust they represent one-third of the total time of the utterance. This datum is in line with Schroeder's findings (2003) where there are some emotions like disgust that are typically expressed by burst, while others such as anger are not. A further observation is that in highly emotionally charged situations, L2 expressive speech is characterized by a very low percentage of disfluencies. L2 speakers rely on spontaneous and relatively universal vocal expressions to transmit their emotional states, instead of editing their performance with self-repairs, repetitions and substitutions. In L1, on the other hand, the percentage of disfluencies, above all in the form of lengthenings and vocalizations, is almost constant in the whole corpus. Filled pauses of this type are typical of spontaneous speech and reflect the speech planning process of the speaker.

5.2. Perceptual data: comparing auto- and external annotations

In order to evaluate the communicative effectiveness of the selected expressive utterances in Chinese-accented Italian, in a second phase of the study the emotional utterances of Chinese speakers were the object of annotation by external Italian and Chinese listeners (henceforth external annotation).

In the external annotation session, since our attention was specifically on the acoustic correlates of emotions in L2, the emotional utterances were administered in the auditory-only mode to 20 Italian and 20 Chinese listeners. They labeled the utterances following the same protocol adopted for the auto-annotation session.

Then, the results of the two labeling methods were compared.

The comparison underlines some similarities and differences between auto- and external annotation and between the rates given by native and non-native listeners. Tables 2 and 3 show the confusion matrices of auto- and external annotation by the two groups of listeners.

		External annotation				
		anger	fear	disgust	anxiety	surprise
Auto-annotation	anger	**29.2**	9.2	3.8	23.1	33.1
	fear	10.3	**48.3**	11.5	16.7	11.1
	disgust	5.8	25.0	**44.2**	7.7	16.3
	anxiety	6.1	14.2	2.4	**61.1**	13.4
	surprise	9.0	10.7	0.6	48.5	**28.6**

Table 2: Confusion matrix of auto- and external annotation by Italian listeners.

		External annotation				
		anger	fear	disgust	anxiety	surprise
Auto-annotation	anger	**26.0**	8.0	4.0	28.0	34.0
	fear	5.6	**42.2**	8.9	8.9	34.4
	disgust	10.0	13.8	**30.0**	13.8	32.5
	anxiety	4.8	14.4	7.4	**66.7**	5.9
	surprise	9.4	9.4	6.1	47.2	**27.8**

Table 3: Confusion matrix of auto- and external annotation by Chinese listeners.

As we can infer from the matrix of table 2, the labels chosen by the native Italians are pretty similar to those selected in the auto-annotation session by the Chinese speakers, particularly in the case of anxiety, fear and disgust. The correspondence between the results of the two annotation methods may be ascribed to the specific prosodic patterns characterizing these emotional states. Anxiety is expressed by the lowest register and the narrowest tonal range. Fear and disgust are transmitted by the same intonation patterns as in L1 Italian, though without such marked differences. Moreover, disgust presents the lowest value of articulation rate and the largest percentage of bursts. The similarities between values of F0 and articulation rate for surprise and anxiety make listeners confuse these two emotions. The emotional state that scores the highest percentage of mismatching between auto- and external annotations is anger, because of the considerable distance of its prosodic pattern from the model produced in L1 Italian. This result can also be explained by considering social conventions and different cultural standards in the expression of

emotional states. As it has been noted in recent literature, Chinese speakers do indeed tend to inhibit the expression of emotions which could threaten relational harmony as in the case of anger (Yang and Campbell 2001; Anolli et al. 2008).

The results of the annotations made by Chinese listeners show that they found it hard to give a specific emotional label to utterances produced in a second language. With the exception of anxiety, recognized by more than 60% of non-native listeners, the rates given by Chinese subjects to the other emotions are definitely more uncertain than those expressed by Italians and more subject to random variations.

6. Conclusions

The objective of the present study was twofold. Firstly we intended to highlight the acoustic check auditory in opening para differences in the expression of the same emotions in L1 and L2. Secondly we aimed to verify the perceptual effects caused by the prosodic patterns of L2 speech on native and non-native listeners.

As for the acoustic correlates of emotions, the spectro-acoustic analyses have demonstrated that L1 and L2 speakers of Italian differ in the way they express emotions. Not only from the lexical, but also from the acoustic viewpoint, the L2 represents the language of emotional distance. Significant variations in the values of the acoustic parameters under study do not always correspond to different emotions. F0 height and tonal range are more reduced in L2 than in L1. The F0 values reached in L1, on the contrary, confirm the tendencies already found in the literature for high and low activation emotions. The data regarding articulation rate show a slowing down in L2, but this parameter does not seem significantly correlated to the five emotional states in L1 nor in L2.

That L2 is the language of emotional distance is particularly evident if we consider speech time composition. The reduced competence in a second language determines a lower percentage of syllable time and an increase of the burst component that occupies 20% and 30% of the utterance in fear, disgust and surprise. During the task, Chinese participants were not allowed to use their L1, the language of emotional expressiveness, and as they could not rely on their second language to express their emotional states, consequently the learners tended to overcome this difficulty by relying on bursts. By burst we mean "very brief, discrete, non-verbal expressions of affection in both face and voice triggered by clearly identifiable events" (Schröder 2003).

As for the perceptual effectiveness of prosodic cues in the expression of emotions in L2 Italian, the data show that there is a strong correlation between the labels assigned by the speakers themselves in the auto-annotation and those chosen by native Italian listeners. This is particularly evident in the case of emotions characterized by similar prosodic patterns in L1 and L2 Italian. The lack of shared prosodic models, on the contrary, provokes a high degree of uncertainty in the listeners when labelling emotional states.

The random judgments given by Chinese external listeners reveal the objective difficulty of L2 learners to identify a specific emotional state in utterances produced in a second language. These considerations lead to the conclusion that Harris' concept of emotional distance is valid both on acoustic and perceptual levels.

A further step for this research will be to extend the Card Task to learners with different mother tongues, and to administer new perceptual tests to native and non-native listeners in order to verify whether the acoustic patterns characterizing the expression of emotions in L2 Italian are imputable to L1 transfer or constrained by interlanguage development.

Bibliography

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Doherty Gwyneth, Garrod Simon, Isard Stephen, Kowtko Jacqueline, McAllister Jan, Jim Miller, Sotillo Catherine, Thompson, Henry S. and Regina Weinert. "The HCRC Map Task Corpus." *Language and Speech*, 34(4) (1991): 351-366.

Anolli, L., Lei Wang, Fabrizia Mantovani, and Alessandro De Toni. "The Voice of Emotion in Chinese and Italian Young Adults." *Journal of Cross-Cultural Psychology*, 39 (2008):565-598.

Aubergé, Véronique, Albert Rilliard, and Nicolas Audibert. "Auto-annotation: an alternative method to label expressive corpora". In *LREC 2006: Workshop on Emotional Corpora*, Geneva, Italy, (2006):45-46.

Belin, Pascal, Shirley Fecteau, and Catherine Be´dard. "Thinking the voice: neural correlates of voice perception." *Trends in Cognitive Sciences* 8(3) (2004): 129-135.

Coan, James A., and John J.B. Allen (edited by). *The Handbook of Emotion Elicitation and Assessment*. Oxford: Oxford University Press, 2007.

Council of Europe. *The Common European Framework of Reference for Languages: Learning, teaching assessment*. Cambridge: Cambridge University Press, 2002.

Dewaele, Jean-Marc. "The Emotional Force of Swearwords and Taboo Words in the Speech of Multilinguals." *Journal of Multilingual and Multicultural Development* 25(2-3) (2003): 204-222.

Dewaele, Jean-Marc. "The emotional weight of I love you in multilinguals' languages." *Journal of Pragmatics* 40(10) (2008): 1753-1780.

Dewaele, Jean-Marc, and Aneta Pavlenko. "Emotion Vocabulary in Interlanguage." *Language Learning* 52(2) (2002): 263–322.

De Meo, Anna and Massimo Pettorino. "L'acquisizione della competenza prosodica in Italiano L2 da parte di studenti sinofoni". In *La didattica dell'Italiano a studenti cinesi e il progetto Marco Polo* edited by Elisabetta Bonvino and Stefano Rastelli, 67-78. Pavia: Pavia University Press, 2011a.

—. "Prosodia e Italiano L2: cinesi, giapponesi e vietnamiti a confronto." In *Apprendere l'Italiano da lingue lontane: prospettiva linguistica, pragmatica, educativa*, edited by Rossella Bozzone Costa, Luisa Fumagalli, and Ada Valentini, 59-72. Perugia: Guerra Edizioni, 2011b.

Ekman, Paul. "An argument for basic emotions." *Cognitive and Emotion* 6 (1992): 169-200.

Hartman, David E., and Jeffrey L. Danahuer. "Perceptual features of speech for males in four perceived age decades." *J. Acoust. Soc. Am*. 59 (1976): 713–715.

Harris, Catherine L., Ayese Aycicegi, and Jean Berko Gleason. "Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language." *Applied Psycholinguistics*, 24 (2003): 561-578.

Harris, Catherine L., Jean Berko Gleason, and Ayese Aycicegi. "When is a First Language More Emotional? Psychophysiological Evidence from Bilingual Speakers." In *Bilingual minds: Emotional experience, expression, and representation*, edited by Aneta Pavlenko, 257-282. Clevedon: Mulilingual Matters, 2006.

Harris, Catherine L. "Bilingual Speakers in the Lab: Psychophysiological Measures of Emotional Reactivity." *Journal of Multilingual and Multicultural Development,* 2(2004): 223-247.

Gerrard-Hesse, Astrid, Kordelia Spies, and Friedrich W. Hesse. "Experimental inductions of emotional states and their effectiveness." *British Journal of Psychology*, 85 (1994):55-78.

Johnstone, Tom, and Klaus R. Scherer. "Vocal communication of emotion." In *Handbook of emotion* (2nd ed.), edited by Michael Lewis, and Jeannette M. Haviland, 220-235. New York: The Guilford Press, 2000.

Jones, Mark, Flora Anagnostou, and Jo Verhoeven. "The vocal expression of emotion: an acoustic analysis of anxiety." In *Proceedings of the 17th International Congress of Phonetic Sciences,* Hong Kong, 17-21 August 201, edited by Wai Sum Lee, and Eric Zee, 982-985, 2011.

Jürgens, Rebecca, Kurt Hammerschmidt, and Julia Fischer. "Authentic and Play-Acted Vocal Emotion Expressions Reveal Acoustic Differences", *Frontiers in Psycholy,* 2 (2011):180. Accessed June, 2014 doi: 10.3389/fpsyg.2011.00180.

Kaneko, Tomoko. "How non-native speakers express anger, surprise, anxiety and grief: a corpus-based comparative study", 2003. Accessed May, 2014
http://ucrel.lancs.ac.uk/publications/CL2003/papers/kaneko.pdf

Karpf, Anne. *The human voice*. London: Bloomsbury, 2006.

Linville, Sue Ellen. "The sound of senescence." *J. Voice* 10 (1996):190–200.

Mullennix John W., Keith Johnson, Meral Topcu-Durgun, and Lynn Farnsworth. "The perceptual representation of voice gender". *J. Acoust. Soc. Am*. 98 (1995): 3080–3095.

Osgod Charles Egerton, William H. May, and Miron S. Murray. *Cross-cultural Universals of Affective Meaning*. Urbana: University of Illinois Press, 1975.

Pellegrino, Elisa. "The perception of foreign accent and speech. Segmental and suprasegmental features affecting degree of foreign accent in Italian L2." In *Proceeding of the VIIth GSCP International Conference – Speech and Corpora*, edited by Heliana Mello, Massimo Pettorino and Tommaso Raso, 261-267. Firenze: Firenze University Press, 2012.

Pettorino, Massimo, Elisa Pellegrino, and Marta Maffia. ""Young" and "Old" Voices: the prosodic auto-transplantation technique for speaker's age recognition". *Speech Prosody*, (2014):135-139.

Pfitzinger, Hartmut R., Noam Amir, Hansjörg Mixdorff, and Jessica Bösel. "Cross-language perception of Hebrew and German authentic emotional speech." In *Proceedings of ICPhS2011*, Hong Kong, 1586-1589, 2011.

Poggi, Isabella, and Emanuela Magno Caldognetto. "Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici". In *Atti del Convegno "Italiano parlato"*, Napoli 14-15 febbraio 2003, edited by Federico Albano

Leoni, Franco Cutugno, Massimo Pettorino and Renata Savy. Napoli: D'Auria Editore, CD-Rom, 2004.

Scherer, Klaus R. "Affect Bursts." In *Emotions*, edited by Stephanie H. M. van Goozen, Nanne E. van de Poll, and Joseph A. Sergeant, 161-193. New Jersey: Lawrence Erlbaum, 1994.

—. "Vocal Communication of Emotion: a Review of Research Paradigm." *Speech Communication*, 40 (2003):227-256.

Schröder, Marc. *Speech and Emotion research. An overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis*, PhD Dissertation, 2003a.

—. "Experimental study of affect bursts." *Speech Communication, Special Issue following the ISCA Workshop on Speech and Emotion*, 40:1-2, 99-116, 2003b.

Schroeder, Marc, Roderick Cowie, Ellen Douglas-Cowie, Machiel Westerdijk, and Stan Gielen. "Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis." In *Proceedings of the EUROSPEECH 2001*, Aalborg, Denmark, 3-7 September, 1:87-90, 2001.

Vidrascu, Laurence, and Laurence Devillers. Annotation and detection of blended emotions in real human-human dialogs recorder in a call center. In *Proc. Affective Computing and Intelligent Interaction* Beijing, 2005.

Yang, Li-Chiung and Nick Campbell. "Linking form to meaning: the expression and recognition of emotions through prosody." In *Proceedings of the 4th ISCA Workshop on Speech Synthesis*, Perthshire, Scotland, 2001.