Simulated Emotional Speech in L2-Pronunciation Training

Federica Missaglia†

† Università Cattolica del Sacro Cuore (Milano)

E-mail: federica.missaglia@unicatt.it

ABSTRACT

This paper aims at verifying to what extent the simulation of different emotions can be employed in L2-German pronunciation training addressed to adult Italian learners. Experimental data on the phonetic characteristics of learners' emotional speech are investigated with regard to specific aspects of the learners' interlanguage. It can be hypothesized that a knowledge of the difficulties native speakers of Italian have to cope with when they learn German and of the phonetic characteristics of their simulated emotional speech, can be combined in L2-pronunciation training in order to promote correct L2 phonetic acquisition and to eliminate the so-called foreign accent.

1 INTRODUCTION

Language-specific and contrastive analyses of German and Italian phonetics and phonology give evidence of great differences at all – segmental, intersegmental and suprasegmental – levels, which can mostly be explained by referring to the distinction between so-called stress-timed and syllable-timed languages.

In line with a 'weak', *i.e.* phonologically oriented version of the stress- *vs.* syllable-timing distinction, with its complex phonotactics and its typical reduction processes depending on prominence ('weak forms'), German can be viewed as a better representative of stress-timing, whereas Italian, characterized by simple syllable structures (mainly CV), no clusters in syllable coda, limited reduction processes and a fairly stable vowel system, tends towards syllable-timing.

The opposite tendencies towards stress- or syllable-timing typically affect not only the speakers' mother tongue but also the learners' interlanguage: While in connected speech native speakers of German tend to reduce and elide segments in unstressed syllables [1], Italian learners typically show the tendency towards elaboration, strengthening of segments and *schwa*-epenthesis [2].

Most phonological interferences in L2-German by Italian learners concerning vowel production [3], accentuation and deaccentuation processes, word and sentence stress assignment and intersegmental co-ordination processes [2], can be attributed to a lack of competence at the suprasegmental level. Empirical evidence showed that in L2-pronunciation training which aims at improving communicative competence rather than formal correctness, conscious attention and awareness for prosodic regularities (a sort of "prosodic awareness") and for emotional-affective aspects in communication is much more effective than memorizing isolated sounds and automatizing abstract articulation patterns [2]. Prosodic awareness – in L1 and L2 – enhances the acquisition of L2 prosodic competence. Once learners acquire a rudimental prosodic competence, many phonological interferences disappear, suggesting that prosody has a controlling function over syllables and segments.

The positive empirical results of prosody-centred pronunciation training compared with segment-centred training in view both of L2 prosody and segments and of the emotional component involved in the acquisition process evidence the need to invert the traditional priorities in L2 pronunciation training and to give prosody a primary role in SLA [4].

Being a sort of interface between the language's grammar and the speaker's intentions, feelings and emotions, prosody can be trained either in relation with its 'grammatical' function, or considering its emotional implications. Speaker's emotions affect neuromuscular co-ordination processes, laryngeal muscle activity, muscle tension and supralaryngeal configurations, thus prosodic features, accentuation and deaccentuation processes, speech rate, articulatory precision and intersegmental co-ordination processes.

It can be hypothesized that the empirical investigation of the phonetic means employed by Italian learners of German to convey intentions, feelings and emotions – even in simulated emotional speech in their mother tongue, which they control more easily and without anxiety, – can offer new insights into their phonetic habits, which may be relevant for teaching practice in SLA in order to help Italian learners to accomplish correct L2-German pronunciation effortlessly.

2 PHONETIC ASPECTS OF SIMULATED EMOTIONAL SPEECH

2.1 SUBJECTS AND MATERIAL

The *corpus* of the investigation consisted of emotionally marked and unmarked speech samples produced by n= 2 non-professional speakers, adult female Italian university students aged 22 years. They produced 7 short sentences

ISBN 1-876346-48-5 © 2003 UAB

without particular emotional emphasis (neutral condition, N) and simulating anger (A), disgust (D), happiness (H), fear (F), sadness (S) and boredom (B).

As the subjects were non-professional speakers, the recordings were randomized, tape-recorded and given n= 9 Italian native speakers for auditive judgements. In separate sessions the judges had to evaluate the sentences' naturalness and to identify the subjects' intended emotion. Only "natural sounding" sentences which were correctly identified by at least 80% of the judges were used for the experimental investigation. Thus the *corpus* of the study consisted of 64 sentences by subjects F and R: 14 (7+7) "neutral" sentences, 9 (3+6) sentences expressing hot anger, 8 (4+4) expressing disgust, 10 (6+4) expressing happiness, 6 (3+3) expressing boredom, 10 (4+6) expressing fear and 7 (4+3) expressing sadness.

2.2 PROCEDURE

The sentences were DAT-recorded with portable Sharp RX-P1H and the digital recordings (12kHz) were phonetically analyzed in detail at the segmental, intersegmental and suprasegmental level both auditorily and acoustically with Kay's *CSL* and *Multi-Speech Mod. 3700 Software for Windows*.

A first investigation at the segmental level was aimed at determining the degree of accuracy in vowel articulation and the deviations from expected tongue position in terms of vowel over- and undershoot. The vowel quality was computed by extracting the frequencies of F1 and F2. The formant frequencies were determined by linear prediction (LPC-order 12, frame length 20ms) and measured at the local maxima of the LPC-spectrum with the so-called peak-picking-method.

At the suprasegmental level, F_0, intensity and duration were analyzed in order to gain indications relating to laryngeal muscle activity, air pression and lengthening of segments. The analysis was performed at two levels – general and particular. In order to find out general tendencies, F_0 and intensity means, standard deviations and ranges were calculated and the duration of emotionally marked sentences was compared with the duration of the corresponding neutral sentences. Statistics were calculated with *SPSS for Windows* 10.0. Due to differences in pauses, length and phonemic content between the 7 sentences, the speech rate in terms of syllables and segments per second was calculated, too.

Then two label files were prepared: one indicating syllable structure and prominence degree (unstressed vs. stressed syllables, word/sentence stress) and another with a narrow transcription following IPA conventions. The phonetic transcriptions of neutral and emotional speech samples were analyzed and compared in order to determine the effects emotions have on intersegmental co-ordination processes (co-articulations, assimilations, reductions, elaborations, epentheses, *etc.*), which are closely related to speech rate, accuracy of articulation, muscular tension and modifications of the neuromuscular co-ordination dependent on different emotional arousal.

2.3 RESULTS

Vowel Quality

As for each emotion the interindividual vocal differences were not more marked than the intraindividual differences, the F1 and F2 means of all vowels were calculated without distinguishing between speakers but only between emotions, and they were displayed in formant charts.

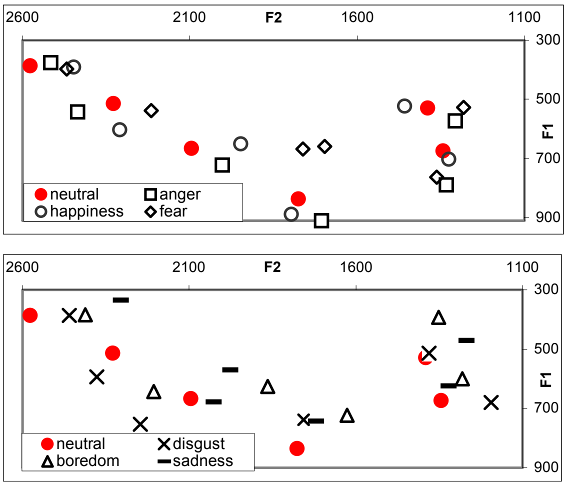

Figure 1: Vowels in neutral and emotional speech.

In vowel production two distinct and opposed tendencies were observed: in the case of happiness and anger, the vowels were either similar to the vowels in neutral sentences (H) or they were articulated in the periphery of the vowel area, becoming more tense and decentralized, thus more distinct (A). In sentences expressing anger the vowel chart became larger: mean F1 was generally higher, mean F2 lower than in neutral condition, indicating vowel overshoot; the vowels were articulated more accurately and with a greater muscular effort than would have been phonologically relevant, *i.e.* than would have been necessary in order to functionally distinguish them.

On the contrary, vowels in sentences expressing sadness and boredom, but also fear, indicated the opposite tendency, *i.e.* towards vowel undershoot: the F1 and F2 means were generally lower indicating a higher and further back articulation than in neutral condition. Front vowels (/i/, /e/, /ɛ/) and (/a/) were generally more centralized and less distinct, indicating that they were articulated with less accuracy than vowels in neutral condition, whereas back vowels (/o/, /ɔ/) became more velarized.

In the case of disgust both tendencies towards vowel over- and undershoot were observed: vowels in extreme positions (/i/ and /a/) were reduced and centralized, whereas medial vowels (/e/, /ɛ/ e /ɔ/) were decentralized.

Prosodic Features

In all emotionally marked sentences, except from the sentences expressing disgust, mean F_0 was higher than in the neutral sentences. The highest F_0 means were observed in the sentences expressing happiness and fear, followed by the sentences expressing sadness, whereas the sentences expressing anger and boredom were articulated with a F_0 which was only slightly higher than the F_0 in neutral sentences.

All emotionally marked sentences were produced with a higher mean intensity than neutral sentences. Mean intensity in the sentences expressing happiness, fear and sadness was higher than in the sentences expressing disgust and boredom.

Simulating happiness and fear both subjects' mean duration was lower than in the corresponding neutral sentences, whereas simulating disgust, boredom and sadness, the mean duration was higher. Only for the sentences expressing anger were interindividual differences observed: the mean duration of F's sentences was higher than the duration of the corrresponding neutral sentences, whereas the mean duration of R's sentences was lower.

Prominence Assignment

It is generally stated that prominence is realized and perceived on the basis of quantitative differences in terms of intensity, duration and F_0, but it is also a well-known fact that there are language-specific differences in the extent to which these three prosodic parameters are effectively employed in prominence assignment. In Italian it has been shown that both in production and perception, duration and intensity are more relevant for prominence assignment than F_0 [5], a finding which is often employed to classify Italian as a syllable-timed language.

In comparing segments' duration and analyzing intensity and F_0 peaks, great divergencies between the emotionally marked sentences were observed, which indicate divergencies in the extent to which the different prosodic features are employed for prominence assignment.

In sentences expressing anger and happiness the curves presented many peaks in intensity and F_0. Statistics also revealed high values for range and standard deviation; the subjects produced many stressed syllables mostly employing F_0 and intensity and not duration for word and sentence stress assignment.

On the contrary, in sentences expressing boredom, disgust, fear and sadness the subjects employed intensity and duration for stress assignment. Especially for disgust, duration affected not only the (vocalic) syllable *nuclei*, but also, and to a great extent, the consonants in syllable onset. Even in the sentences with great fluctuations in the F_0 curve, as in the case of disgust and sadness, F_0 did not have a phonological function for stress marking.

Speech rate

The data concerning speech rate confirm the existence of two opposite tendencies in simulated emotional speech: while the sentences simulating disgust, boredom and sadness were produced at a lower speech rate, those simulating anger, happiness and fear were produced at a higher speech rate than the corresponding sentences in neutral condition. This relation between slower and faster articulation emerged both computing the speech rate in terms of syllables per second (A: 6.112988 sill./s, H: 6.120203 sill./s, F: 6.138204 sill./s *vs.* D: 5.344147 sill./s, B: 5.7981 sill./s, S: 5.834915 sill./s) and of segments per second (A: 13.35801 seg./s, H: 13.44438 seg./s, F: 12.98079 seg./s *vs.* D: 10.90642 seg./s, S: 12.43758 seg./s, B: 12.61939 seg./s).

Intersegmental Co-ordination Processes

The intersegmental co-ordination processes were classified as elaborations (epenthesis, diphthongation, decentralization, aspiration, affrication) or reductions (elision, monophthongation, centralization, sonorization, desonorization, assimilation, nasalization), and the differences between the intersegmental co-ordination processes in neutral and emotional speech were analyzed both in terms of quantity and quality.

Data on the number of segment elisions and epentheses with respect to neutral speech (for the formula see [6]) confirm the existence of two tendencies: fewer segmental elisions were observed in sentences expressing anger and happiness, more elisions in sentences expressing disgust and boredom, whereas fear and sadness showed results at an intermediate level.

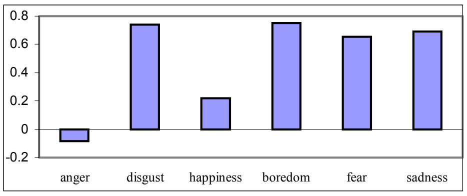

Figure 2: Number of elided/inserted segments (positive values indicate segment elisions, negative values indicate segment epentheses).

The analysis of the phonetic transcriptions revealed that the sentences expressing fear and anger were the emotions most affected by intersegmental co-ordination processes, both by elaborations and by reductions. For the other emotions a distinction between the number of elaborations and of reductions has to be drawn: while boredom and disgust were less affected by elaborations than sadness and happiness, sadness and happiness were less affected by reductions than boredom and disgust. Except for boredom and sadness, generally intersegmental co-ordination processes affected vowels more than consonants.

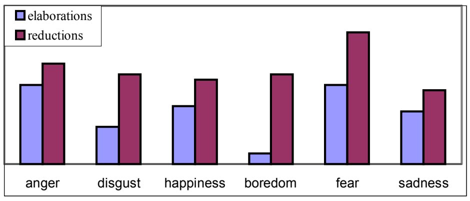

Figure 3: Intersegmental co-ordination processes.

3 DISCUSSION

The auditory and acoustic analyses of Italian learners' emotional speech reveal great phonetic divergencies between the simulation of different emotions, which can be put in relation with their degree of neurophysiological arousal (high arousal for anger, happiness and fear *vs.* low arousal for disgust, boredom and sadness). The degree of emotional arousal influences the laringeal muscle activity, responsible for prosodic features, and it also affects speech rate, muscle tension and supralaringeal configurations, which are responsible for the precision of segmental articulation and the intersegmental co-ordination processes.

Emotions with high neurophysiological arousal were generally characterized by greater muscular constriction and articulatory effort, which increased not only mean F_0 and intensity, but also the accuracy of articulation; vowel overshoot and segment elaborations were frequent. In spite of a higher speech rate, in emotions with high arousal the articulation of single segments was more accurate than in emotions with low arousal and a lower speech rate. This apparent contradiction can mostly be explained with regard to prosodic features, as emotions with high arousal were generally characterized by marked prosody (many F_0 and intensity peaks, many stressed syllables). Here prominence was mostly assigned by variations in F_0 and intensity.

Emotions with low arousal showed the opposite tendency, due to reduced motor activity and muscular tension: vowels were more centralized and reduced, consonants were articulated with less accuracy; vowel undershoot and consonant reduction processes such as assimilations, coarticulations and elisions were frequent. In spite of a lower speech rate, the high number of unstressed syllables which characterized especially the sentences expressing disgust and boredom, had as an effect that they were articulated with a low muscular effort and with less accurate and precise articulatory gestures, thus generating many reduction processes. The sentences were generally characterized by homogeneous F_0 curves; F_0 had a secondary role in prominence assignment with respect to intensity and especially duration both of vowels and of consonants.

These findings reveal that one group of emotions is particularly apt for use in L2-German pronunciation training addressed to Italian learners, *i.e.* emotions with low neurophysiological arousal, because when simulating disgust, boredom and sadness, learners spontaneously attain phonetic break-throughs of the phonological characteristics of syllable-timing and realize typically German phonetic reductions (vowel centralizations, segment elisions, *etc.*) which normally cause great difficulties in their German interlanguage. On the contrary, when simulating emotions with high neurophysiological arousal, such as anger, happiness and fear, they preserve the phonological characteristics of their mother tongue which are related to syllable-timing and even reinforce their natural and typically Italian tendencies towards phonetic elaborations, *schwa*-epenthesis and the realization of many stressed syllables. Even if the speech rate is faster, they spontaneously simplify the phonotactic construction of all syllables to the same extent by reducing the number of segments both of stressed and unstressed syllables.

4 CONCLUSION

The empirical data on the phonetic nature of simulated emotional speech in learners' L1 in terms of segmental reductions, articulatory settings, prosodic features and intersegmental co-ordination processes offer important findings which can be applied in L2-German pronunciation training. As a next step it is necessary to verify experimentally whether the spontaneous tendencies in L1 simulated emotional speech are preserved in L2 simulated emotional speech as well.

AKNOWLEDGEMENTS

This work was supported by Università Cattolica del Sacro Cuore, Linea di Ricerca D1-2002.

REFERENCES

[1] K. Kohler, "Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations," in *Speech Production and Speech Modelling*, W.J. Hardcastle and A. Marchal, Eds., pp. 69-92. Dordrecht: Kluwer, 1990.

[2] F. Missaglia, *Phonetische Aspekte des Erwerbs von Deutsch als Fremdsprache durch italienische Muttersprachler*, Frankfurt a.M.: Hector, 1999.

[3] F. Missaglia and W.F. Sendlmeier, "Die Realisierung deutscher Vokale durch italienische Muttersprachler – Eine experimentalphonetische Untersuchung," *Zeitschrift für Fremdsprachenforschung*, vol. 10/1, pp. 73-95, 1999.

[4] F. Missaglia, "Contrastive Prosody in SLA: An Empirical Study with Italian Learners of German," *Proceedings of the XIV ICPhS,* San Francisco, vol. 1, pp. 551-554, 1999.

[5] P.M. Bertinetto, *Strutture prosodiche dell'italiano. Accento, Quantità, Sillaba, Giuntura, Fondamenti metrici*, Firenze: presso l'Accademia della Crusca, 1981.

[6] M. Kienast and W.F. Sendlmeier, "Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech," in *Speech and Signals*, W.F. Sendlmeier, Ed., Frankfurt a.M.: Hector, pp. 157-168, 2000.