

ARTICULATORY REDUCTION IN EMOTIONAL SPEECH

Miriam Kienast, Astrid Paeschke, Walter Sendlmeier
Technical University Berlin, Institute of Communication Science, Germany
kienast@kgw.tu-berlin.de

ABSTRACT

The present study aims at examining vocal expression of emotion. Emotionally loaded speech material produced by actors was analyzed with reference to the accuracy of articulation as well as to the duration of syllables and segments. It was investigated whether the vocal expression of several emotions (anger, happiness, fear and sadness) differ from one another and from neutral speech with respect to the durational and qualitative reduction of articulation. The obtained results suggest that accuracy of articulation and durational aspects are useful parameters to characterize especially the emotions anger, fear and sadness.

1. INTRODUCTION

Many studies of the last 50 years concerned with the vocal expression of emotion dealt with the investigation of prosodic and durational parameters as well as intensity. The inquiry of articulation and especially the accuracy of articulation did not rank first. So it seems reasonable to take this parameter into consideration, because accuracy of articulation is closely related to two other parameters, namely speech rate and muscular tension [1] influenced by different arousal of the autonomic nervous system. And it is a widely known fact that several emotions differ with respect to these parameters [2, 3, 4]. The few studies, which examined changes of articulation in emotional speech, aimed at the investigation of this parameter predominantly by examining formants. E. g. Goldbeck et al. [5] compared the real formant position with an ideal one to measure the deviation of the tongue from the aspired position. In the present study a different approach was used. Here the accuracy of consonant articulation was focused by analyzing the number and extent of segmental reductions on the basis of Kohler's concept of reduction forms in German [1,6]. This concept describes the deletion and assimilation of segments as a result of different principles, namely reduction of effort, listener orientation, gestural reorganization, cognitive constraints, and speaking styles.

Beside this examination of the qualitative reduction of articulation the second objective of this study was to examine the quantitative segmental changes. Therefore the question was addressed, how the alteration of the speech rate affects the duration of single segments in

detail and if there are differences between the several classes of consonants.

2. MATERIAL

The database used in this examination consists of utterances which were produced by one male and two female German actors enacting the emotions anger (hot), happiness, fear (as a kind of panic), sadness and a neutral version. Because of the differences between read speech and spontaneous speech concerning articulatory reduction as noted by Kohler [1], attention was paid that the actors produce the speech material in a spontaneous manner. The applied utterances are four sentences with neutral semantic content useable in everyday communication. The occurrence of articulatory reduction is influenced by the articulatory context of each segment. So the phonotactic construction of the sentences provides the possibility of all reduction forms listed by Kohler [1]. The emotional content of the speech material and its naturalness were evaluated in a perception experiment. The recorded sentences were randomized and acoustically presented to 20 naive listeners. Only those utterances which were judged as natural and for which the emotion was recognized by at least 80% of the listeners were used for further analysis.

3. METHODS AND PARAMETERS

For all selected sentences two label files were created using spectrographic and oszillographic representations as well as auditory analysis. The first represents the segmental structure and comprises a narrow transcription of the utterance. The second includes the syllable structure of the sentence and a categorization of each syllable into unstressed, word stressed and phrase stressed. Each label file was checked by a second phonetician.

On the basis of this labelling the following reduction forms more detailed described by Kohler [1,6] were analyzed:

- progressive and regressive assimilations with respect to place of articulation and manner of articulation
- assimilations with respect to nasality and voicing
- changes of opening degree
- reorganizations

For each sentence the total number of these reductions was counted and the degree of each reduction classified in either complete (the whole segment was reduced) or partial (only a part of the segment was reduced) was determined. In addition, elaborations were detected as parts of the utterance pronounced more accurately. In order to measure omissions of segments the Lautminderungsquotient (LMQ) developed by Hildebrandt [7] was calculated according to the following equation:

$$LMQ = 10 - \frac{10 \cdot n_{eff}}{n_{fict}}$$

The LMQ provides the possibility to compare the number of the factually produced segments (n_{eff}) with the number of segments which should have been produced for standard pronunciation (n_{fict}). Thus positive values represent segment deletions, while negative values indicate segment insertions. Furthermore the number of the deleted syllables as one of the highest degrees of reduction was determined.

In order to characterize the changes in the time domain the following parameters were measured:

- duration of the whole utterance without pauses
- speech rate as syllables per second
- number and duration of unfilled pauses
- duration of nasals, stops and fricatives, the latter both with respect to voiced and voiceless ones
- duration of the voice-onset-time for voiced and voiceless stops
- duration of vowels separated to the three different stress-levels mentioned above

The values of all parameters for all different emotions were compared with each other and with the values of the neutral version.

4. RESULTS

4.1. Accuracy of Articulation

The results obtained by analyzing the appearing reduction forms and by evaluating the LMQ, revealed the following:

Compared with the neutral speaking style articulatory reduction is more frequent and more extreme in utterances expressing fear and sadness. The number of deleted segments measured by the LMQ is higher for sadness and fear. Figure 1 illustrates that in anxious utterances there are about 3,5 tenth parts of segments more deleted than in the neutral version; in sad utterances 2,5 tenth parts more are deleted. Furthermore about 6% of all syllables were deleted in sad utterances and 4,5% in sentences representing fear, as can be seen in Figure 2.

In contrast, sentences representing anger showed a more accurate pronunciation. Deletion of segments (LMQ) was observed more seldomly than in all other emotions including neutral speech (Figure 1) and the omission of a

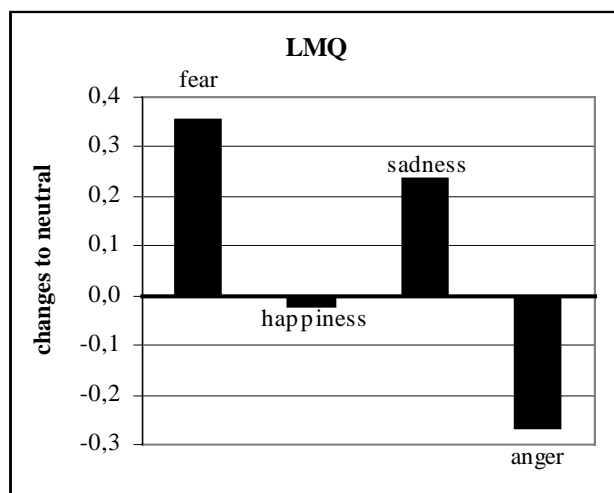


Figure 1. Average LMQ (tenth parts of deleted segments related to neutral speech)

syllable appeared only one time in the whole database. Though the numerical proportion of occurred assimilations was not less than in the neutral versions, the degree of the reduction features is less extreme: Overall for the sentences representing anger the rate of complete assimilations is lower than of partial assimilations, in contrast to the other emotional loaded sentences including neutral speech where this proportion is nearly the

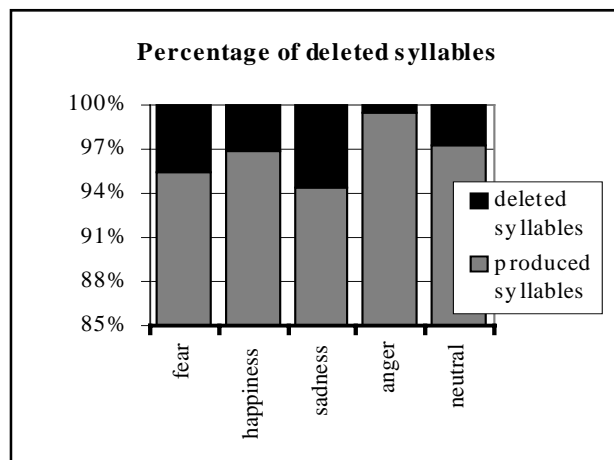


Figure 2. Average percentage of deleted syllables

same. Furthermore anger was the only emotion which provides examples of elaboration.

The results of the utterances expressing happiness differ not significantly from the neutral version. The outcomes revealed that the number and extent of assimilations as well as segment and syllable deletion in the examined sentences with happy content corresponds to the observed accuracy of articulation in neutral speaking.

4.2. Time Domain

4.2.1. Speech rate, total duration and pauses

The results obtained by the analysis of the speech rate, the duration of the whole utterance and pauses confirm widely the results of Scherer [2] and others [3, 4]. Compared with the neutral version and the other

emotions, the sentences expressing sadness are longer and produced with the slowest speech rate (Figure 3) as well as with the highest frequency and with the maximal duration of pauses.

Utterances representing fear were normally short and were produced with a high speech rate (Figure 3). Pauses in this emotion occurred much more seldomly than in

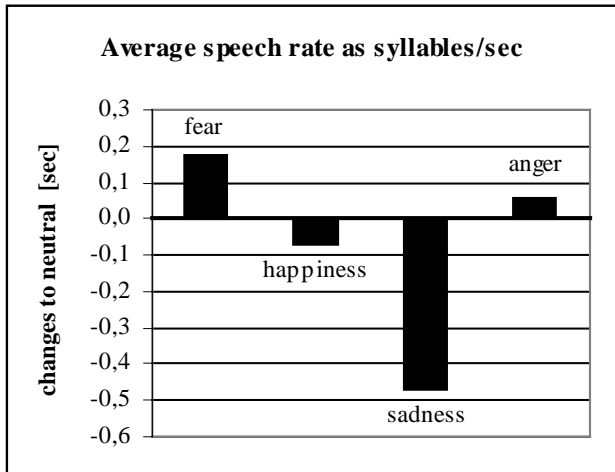


Figure 3. Average speech rate related to neutral speech

utterances expressing sadness and are much shorter, too. The speech rate of sentences expressing anger is a little faster than that of the neutral version (Figure 3), but these differences are not significant. The same applies for happiness which was produced more slowly (Figure 3) than neutral speech. Pauses in speech with happy or angry content are rare in our database and only of short duration.

4.2.2. Segment duration

The further analysis of the durational features of emotional speech suggests for utterances expressing sadness corresponding with the slow speech rate a lengthening of all different classes of segments, namely vowels, nasals, fricatives and stops. Furthermore the durational increase of consonants is higher than that of vowels. So the average increase of consonant duration amounts to 0,3 s more than in neutral speech, while average duration of vowels is only 0,02 s longer than in the neutral version. With regard to the vowel duration in sentences with sad content it was observed that phrase stressed vowels are more prolonged than unstressed and word stressed ones.

For the speech material with angry content the obtained results show a prolongation of vowels, especially word stressed ones, whereas nasals, fricatives and stops are shortened. Figure 4 illustrates the values of the word stressed vowel duration.

For the sentences expressing fear it was found that in conformity with the fast speech rate all kinds of segments are normally shorter than in neutral speech with the exception of voiceless stops. Voiceless stops in anxious utterances are longer than in the neutral version, because they are often strongly aspirated. This also

applies to sad utterances and could be supported by the values of the voice-onset-time-duration shown in Figure 5 which is quite longer for sentences representing fear and sadness.

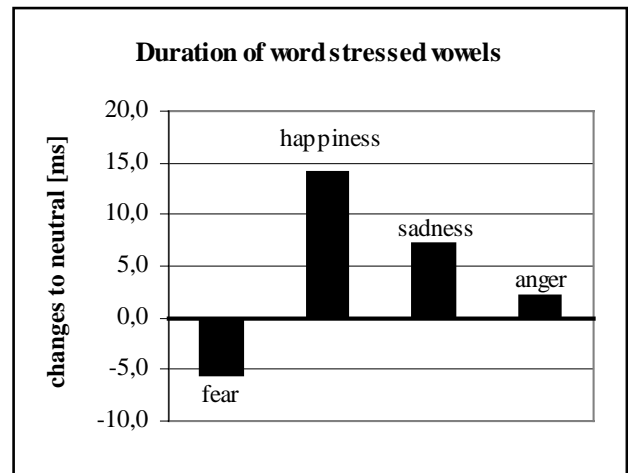


Figure 4. Average duration of word stressed vowels related to neutral speech

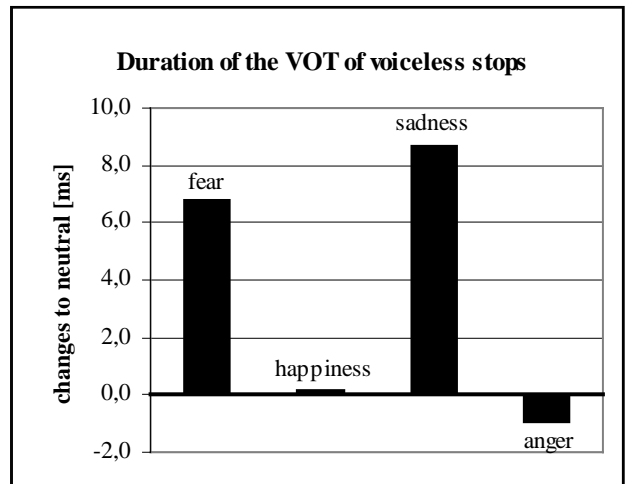


Figure 5. Average VOT-duration of voiceless stops related to neutral speech

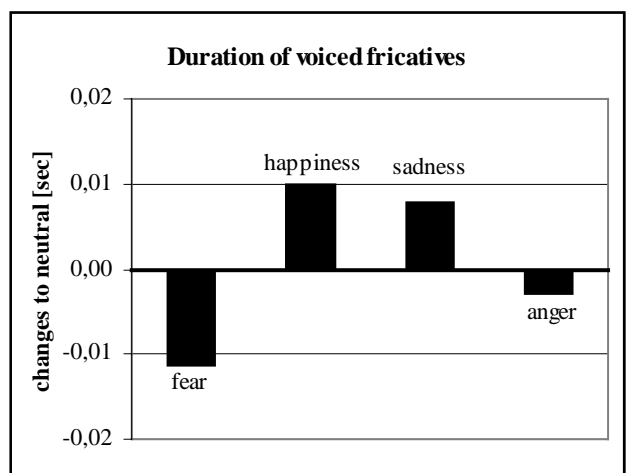


Figure 6. Average duration of voiced fricatives related to neutral speech

The outcomes of the examined sentences with happy content revealed that durational features are more speaker dependent than in the other emotions. However for two classes of segments systematical results were extracted: Voiced fricatives are prolonged in utterances expressing happiness as illustrated in Figure 6. Furthermore phrase stressed vowels also show an increase of duration compared with the neutral version and the other emotions.

5. SUMMARY

In the present study it was observed that the vocal expression of several emotions show different peculiarities with respect to qualitative and durational reduction, especially anger, fear and sadness. The examined utterances expressing anger showed a high accuracy of articulation, shorter consonants and prolonged vowels. These outcomes correspond with other findings [2,3] which noted more distinctive opening and closing movements for anger. Furthermore the results suggest that articulatory reduction is most frequent in the utterances expressing fear and sadness.

This seems reasonable for fear, as it is produced with the fastest speech rate which is one of the most important parameters for reduction contents [1,6]. But for sadness this is remarkable, because here the speech rate is very slow. Some explanation for the high degree of articulatory simplification in sad utterances may be the low muscular tension as a result of the high parasympathical arousal associated with sadness. Durational analysis of utterances expressing sadness showed that all segments are lengthened, especially consonants. Furthermore the continuation of speaking in sad utterances is disrupted by long and frequent pauses. Anxious sentences show shortened segments with exception of voiceless stops which are often aspirated like in speech with sad content. The results of happiness with regard to the segment duration revealed an lengthening of voiced fricatives and phrase stressed vowels, while the accuracy of articulation in utterances expressing happiness seems to be like that in neutral speech.

Since the result of this examination are based on a small sample size, caution should be taken when making any generalized conclusions. Therefore the described analysis will be expanded to more speakers and sentences. Furthermore future work on this issue may take into consideration a more direct gathering of articulatory data in emotional speech, e.g. by articulographic or palatographic measurements.

ACKNOWLEDGEMENT

This work was supported by the DFG (German research foundation, Se 462/3-1)

REFERENCES

- [1] Kohler, K.J. (1995): *Articulatory Reduction in Different Speaking Styles*. Proceedings ICPHS 95, Stockholm, Vol. 2, p. 12-19
- [2] Scherer, K.R. (1986): *Vocal Affect Expression: A Review and a Model for Future Research*. Psychological Bulletin, Vol. 99, No. 2, p. 143-165
- [3] Murray, I.R. & Arnott, J.L. (1993): *Towards a simulation of emotion in synthetic speech: a review of the literature on human vocal emotion*. JASA 93 (2), p. 1097-1108
- [4] Williams, C.E. & Stevens, K.N. (1981): *Vocal correlates of emotional states*. In: Darb, J.K. (ed.): *The evaluation of speech in psychiatry*. New York, Grune & Stratton
- [5] Goldbeck, T., Tolkmitt, F. & Scherer, K.R. (1988): *Experimental studies on vocal content communication*. In: K.R. Scherer (ed.): *Facets of emotion*. Recent Research. Hillsdale: Lawrence Erlbaum
- [6] Kohler, K. (1990): *Segmental reduction in connected speech in German: phonological facts and phonetic explanations*. In: Hardcastle, W.J. & Marchal, A. (eds.): *Speech production and speech modelling*, p. 69-92
- [7] Hildebrandt, B. (1963): *Die arithmetische Bestimmung der durativen Funktion. Eine neue Methode der Lautdauerbewertung*. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 14, p. 328-336