ACOUSTIC-PHONETIC CHARACTERISTICS OF HYPERARTICULATED SPEECH FOR DIFFERENT SPEAKING STYLES

Stefanie Köster

Institute for Communications Acoustics
Ruhr University Bochum (Germany)
koester@ika.ruhr-uni-bochum.de

ABSTRACT

This study aims to describe differences between hyperarticulated and normal speech. Hyperarticulated, or clear speech is produced when addressing to hearing-impaired listeners. It also appears quite often in spoken language systems as the user's reaction on previous recognition errors. In this paper we present a comparison of the acoustic-phonetic characteristics of normal and hyperarticulated speech for three different types of utterances, single words, single sentences and spontaneous speech. Duration, fundamental freqency, formants and formant bandwidths change significantely. Significant differences between the three speaking styles are observable, especially for spontaneous speech vs. words and sentences. We report on an auditory test investigating the percieved changes in the two speech types.

1. INTRODUCTION

An advantage of using spoken language systems for human-machine communication is the fact that the user can enter information fast and naturally to the system, especially in 'hands-busy-eyes-busy' situations, e.g. navigation systems for cars. Since the user of those systems always has to deal with recognition errors, even more in noisy environments, one strategy of overcoming those failures is speaking more clearly and accentuated. Previous studies [1,2] showed that hyperarticulation leads to an increase of intelligibility for human-to-human communication, whereas for an automatic speech recognition system recognition error rates rise [3,4]. Lately, some approaches to improve speech recognition systems with regard to hyperaticulated speech have been established [3,5]. However, most models are based on the analysis of isolated words. Junqua [6] reported differences in acoustic-phonetic features of Lombard speech for various speaking situations as well as for the different genders. If this is also true for hyperarticulated speech, there is a need for more flexible adaptation methods.

This paper is organized as follows: In the first section the database with normal and hyperarticulated speech is described. This includes the contents of the database and the method of producing hyperarticulated speech. Section two describes the results of the statistical analysis of the database. Interesting differences between the three speaking styles are pointed out. After that we present the results of an auditory test, investigating the perceptional differences between normal and hyperarticulated speech. Changes in the percieved speech signal are of interest for the modelling of the speaking style in speech synthesis. We end with a summation of the results given in this study.

2. THE DATABASE

The collected database consists of german normal and hyperarticulated speech. Three different types of utterances were recorded for each speaker. Isolated words contained numbers and typical instructions for robots. The recorded sentences were phonetically balanced. The spontaneous speech was produced by simulating a train reservation system. Special sheets, which were developed for the assessment of telephone line quality [7], were used. Dialogue sheets were designed to produce the same amount of speech for each partner. The recordings took place in the institute's anechoic chamber.

The data was recorded in two sessions. The first scenario was the recording with normal speaking style. Speakers had a person in front of them they could address to. There was a pause between the recordings of the different utterance styles. For the other scenario the second person wore headphones in order to signalize a disturbed communication situation to the talker. Speakers were instructed to talk clearly to the second person, who pretended the disunderstanding of the utterances. There were 3 female and 3 male talkers, each uttering 26 words, 10 sentences and a reservation dialogue.

The recorded utterances were segmented into phones, each annotated with information about speaker, speaker's gender, speaking style, style of utterance and phonetic information.

3. ACOUSTIC-PHONETIC ANALYSIS

For most of the analyzed features of hyperarticulated speech, there are significant changes observable.

3.1 Duration

The general word duration of the data for the three text styles was calculated by eliminating pauses in the speech signal. Duration increases for all three text styles. Figure 1 shows the amount of increase. As you can see, the change of word duration is much higher for sentences than for the dialogue or words.

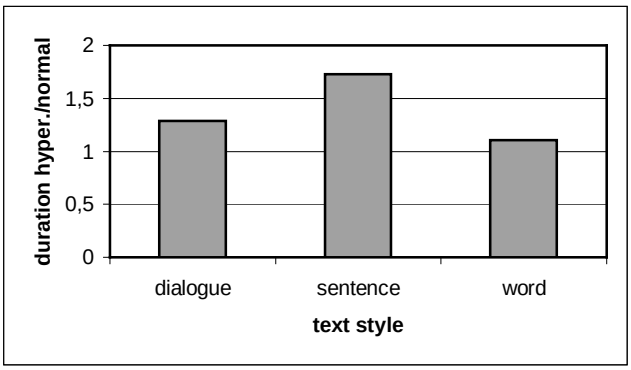

Figure 1: Ratio of duration hyperarticulated/normal speech

Average phone duration increases for all utterance types significantely according to the results of a t-test. For spontaneous speech the amount of increase (6,7%) is far less than for words (12%) and sentences (11,5%). Table 1 shows the changes of average segmental duration for the various phoneme classes. Highlited percentages mark a significant difference. You can see significant changes of more classes of phonemes for spontaneous speech than for words or sentences. Even though it is not significant, there is quite a difference between the average duration of plosives, nasals and liquids of the dialogues vs. those of words and sentences.

	Plosives	Nasals	Liquids	Short vowels	Long vowels	Schwas
Dialogue	-11,6	-15,1	**32,5**	**17,5**	**20,5**	**25,7**
Sentence	1,3	11,6	7,3	**24,0**	**20,5**	9,5
Word	15,1	3,9	-5,2	**26,7**	**15,6**	14,3

Table 1 : Percentages of changes in average segmental duration

As you can see in Table 1, the most important change in phoneme duration happens for vowels. Figure 2 shows the average change of vowel duration for different syllable positions in the sentence for dialogue and sentences. For spontaneous speech, the lengthening of vowels mostly takes place at the beginning and the end of sentences, while for read sentences mostly the vowels inbetween the phrase boundaries are affected. Since the lengthening indicates an emphasizing of syllables, it seems to be more important to emphasize the beginning and ending of a sentence in spontaneous speech. For single sentences the beginning and end is clear to the speaker, so every syllable gets the same emphasize.

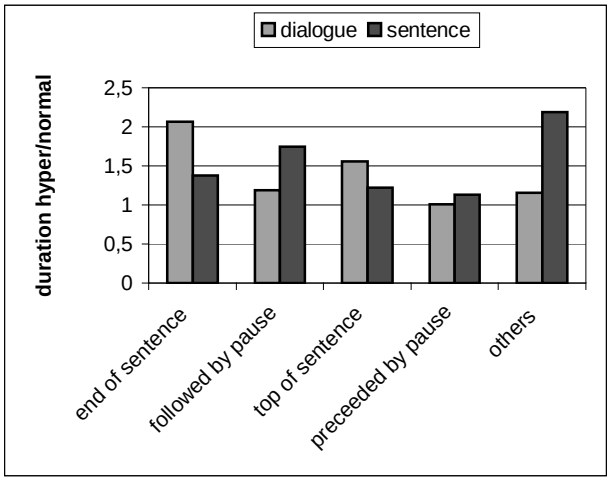

Figure 2: -Ratio of vowel duration for hyperarticulated / normal speech for different syllable positions

3.2 Fundamental Frequency

For all three speaking styles, fundamental frequency increases. The biggest difference is observable for words (25,0%). Changes for spontaneous speech and sentences are almost the same (21,5% and 21,2%). Figure 1 shows a boxplot of mean F0 for all phonemes. You can see an increase of the variation for spontaneous speech, which is not observable for sentences and words.

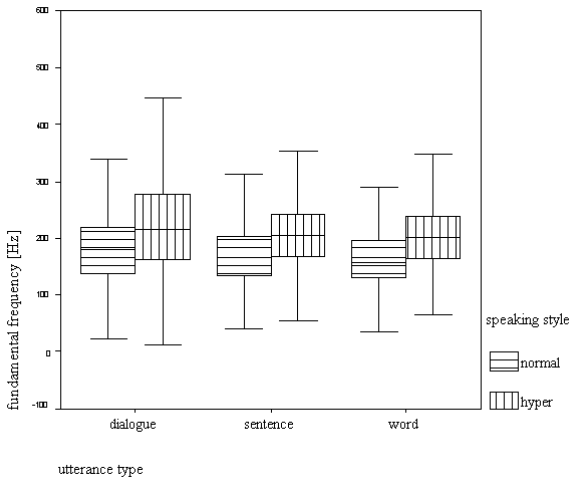

Figure 3: Average F0 in Hertz for the three utterance types

3.3 Formants and Formant Bandwidths

Only for sentences there is a slight increase of formant frequencies observable. Table 2 shows the amount of change in percentages. For all text styles formant bandwidths are much lower for hyperarticulated speech

than for conversational speech. All changes are significant (t-test).

	F1	F1 band	F2	F2 band	F3	F3 Band
Dialogue	-14,2	-27,2	-19,0	-37,2	-12,8	-39,8
Sentence	2,9	-13,1	1,9	-9,2	-0,2	-7,9
.Word	-5,7	-54,9	-6,5	-54,9	-0,8	-50,9

Table 2: Average of formant and formant bandwidth frequency changes of all voiced phonemes in percent

There are only small differences between the various phoneme classes. For all classes the formant frequencies decrease for hyperarticulated speech, especially for fricatives. Figure 3 shows the F1-F2 plane for long vowels of hyperarticulated and normal speech. You can clearly see the shift towards lower frequencies for hyperarticulated speech. This result is not the same as Pitcheny[] gets for English hyperarticulated speech, as he could not observe a big difference between the speaking styles for English.

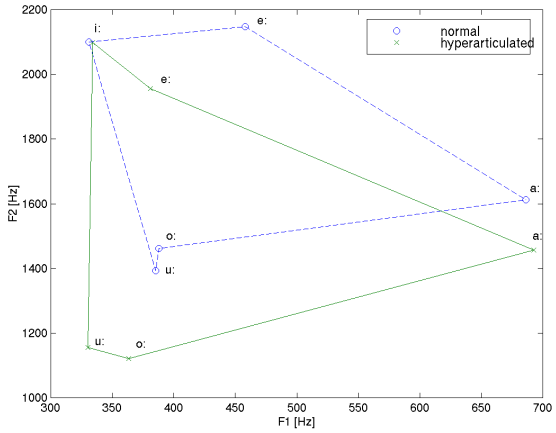

Figure 4: Formant frequencies in F1-F2 plane for long vowels

Pitcheny analyzed only sentences. The shift here results from the data of spontaneous speech rather than that of sentences or words. For the latter two utterance types, there are only small changes for long vowels observable (data not shown here), which confirms Pitcheny's observations.

3.4 Spectral Tilt

The spectral tilt of plosives, fricatives and affricatives tends to be flatter for hyperarticulated speech than for normal speech, while for the other phoneme classes, especially for vowels, spectral tilt becomes steeper.

4. PERCEPTIONAL DIFFERENCES

An auditory test was performed in order to find the perceptional differences between hyperarticulated and normal speech.

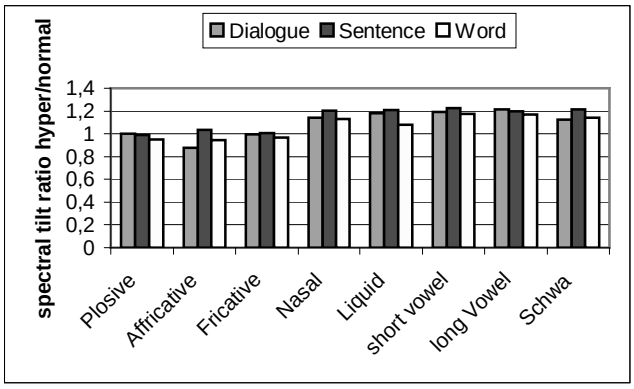

Figure 5: spectral tilt ratio of hyperarticulated/normal speech for different utterance types

4.1 Test Setup

The idea was to let the subjects describe the utterances by a set of antonymes. The set was low-high, comfortable-uncomfortable, dark-light, soft-solid, slow-fast, expressive-expressionless, reduced-hyperarticulated, monotonous-melodic, informative-appellative, restricted-aggressive, clear-nasalized, powerless-powerful and formal-spontaneous. Nineteen subjects listened to 14 sentences of normal and 14 sentences of hyperarticulated speech. For each stimulus they had to decide on each pair of antonymes by describing a value between 0 (p.e. slow) and 6 (p.e. fast). Stimuli were heard via headphones. The subjects were allowed to repeat stimuli for judgement.

4.2 Results

Figure 6 shows the results of the auditory test for all antonyme pairs. Hyperarticulated speech is judged as fast or slow as normal speech. This is quite extraordinary as the over all duration of the hyperarticulated speech signals is larger than for normal speech signals (section 3.2). Another remarkable result is that hyperarticulated speech is not judged as extremely hyperarticulated. But it is percieved as much more uncomfortable. A reason for this could be the unawareness of the subjects of what hyperarticulation really is. An interesting judgement is that hyperarticulated speech is much more aggressive and powerful than normal speech.

A clustering of the judgements hase been performed. A third of the stimuli were clustered as hyperarticulated speech, while two thirds were clustered as normal speech. This could mean, that not all of the hyperarticulated stimuli have all of the typical attributes of that speaking style. Table 4 shows the centers of the cluster 'hyperarticulation'. An average of three means a neutral result for the pair of antonymes.

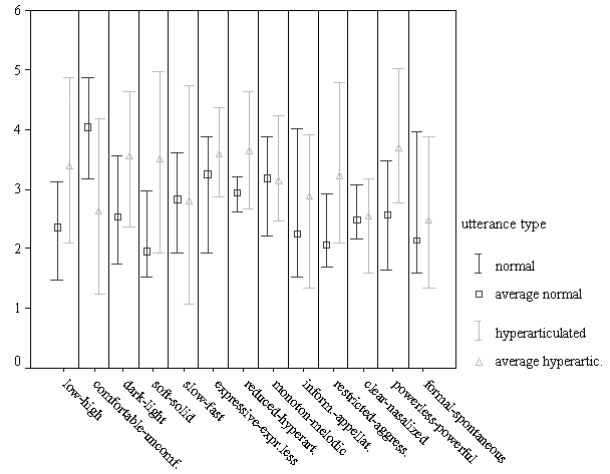

Figure 5: Minimum, maximum and average judgements for antonymes

A factor analysis showed that there are four prominent independent dimensions; Dimension 1: restricted-aggressive, soft-solid and powerful-powerless; Dimension 2: expressive-expressionless and monotonous-melodic; Dimension 3: formal-spontaneous, reduced-hyperarticulated and informative-appellative; Dimension 4: low-high, slow-fast, clear-nasalized and dark-light. In dimension 4 you can see that the attributes related to physical quantities are not percieved independently from each other. The only excepetion are the attributes monotonous-melodic, which can be related to the variance of fundamental frequency.

	Cluster	
	normal	hyper
low-high	2,36	3,86
comfortable-uncomfort.	4,01	1,981
dark-light	2,62	3,83
soft-solid	1,90	4,37
slow-fast	2,57	3,39
expressive-expr.less	3,24	3,72
reduced-hyperart.	2,94	3,93
monotonous-melodic	3,18	3,04
informative-appellative	2,16	3,41
restricted-aggressive	1,89	4,11
clear-nasalized	2,46	2,53
powerless-powerful	2,42	4,43
formal-spontaneous	2,05	2,88

Table 3: Centers of clustering

5. CONCLUSION

We have described the recordings, the acoustic-phonetic characteristics and the perception of hyperarticulated and normal speech. We compared three different utterance types, spontaneous speech, single sentences and single words. We found important differences among the acoustic-phonetic characteristics of the text types. The impression of hyperarticulated speech subjects got from an auditory test were quite surprisingly. Hyperarticulated speech was judged as powerful and aggressive. An interesting question is the effects of hyperarticulation on intelligibility for spontaneous speech for human speech recognition, especially in comparison to Lombard speech.

6. ACKNOWLEDGEMENTS

The author would like to thank Christine Dudda, Milena Puzig and Julia Hucke for their support. Thanks also to the staff at the Institute for Communication Acoustics of Ruhr-University Bochum and Prof. Jens Blauert.

7. REFERENCES

[1] M. Picheny, D. Durlach, and L. Braida. Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. In *Journal of Speech and Hearing Research*, vol. 28, March 1985, pp. 96-103

[2] K. Payton, R. Uchanski, and L. Braida. Intelligibility of conversational and clear speech in noise and reveberation for listeners with normal and impaired hearing. In *Journal of the Acoustical Society of America*, **95** (3), March 1994, pp. 1581-1592

[3] S. Oviatt, M. MacEachern, and G. Levow. Predicting hyperarticulate speech during human-computer error resolution. In *Speech Communication*, **24** (2), 1998, pp. 1-23

[4] H. Soltau, and A. Weibel. On the influence of hyperarticulated speech on the recognition performance. In *Proceedings of the International Conference on Spoken Language Processing*, Sydney, Australia, 1998

[5] H. Soltau, and A. Waibel. Specialized acoustic models for hyperarticulated speech. In *Processings of the International Conference on Spoken Language Processing*, Istanbul, Turkey, 2000

[6] J. Junqua, S. Fincke, and K. Field. The Lombard effect: a reflex to better communicate with others in noise. In *Proceedings of the International Conference on Spoken Language Processing*, Phoenix, USA, 1999

[7] ITU-T Contribution COM 12-35. Development of scenarios for a short conversation test. Source: Federal Republic of Germany (S. Möller). International Telecommunication Union, CH-Geneva, 1997