

Rate dependent spectral reduction for voiceless fricatives

Benjamin Weiss

Quality & Usability Lab, Deutsche Telekom Laboratories, Berlin University of Technology

BWeiss@telekom.de

Abstract

The relationship between local speaking rate and spectral reduction of voiceless fricatives is examined based on a corpus of spontaneous speech. The spectral balance of about 18k items is statistically analyzed with regard to consonant duration and local speaking rate. Both measures show comparable effects: For most of the treated allophones, significant within-speaker reduction occurs with higher rates. The speaking style of the material is constant and stress, part of speech, and position in the syllable was controlled. However, as the observed effects are small compared to differences between allophones, perceptual relevance is considered to be implausible.

Index Terms: speaking rate, center of gravity

1. Introduction

In contrast to vowels, spectral consequences of articulatory reduction of consonants are investigated only rarely. However, results of those acoustical studies reveal that consonants produced with a more informal style or in unstressed condition are reduced in duration and spectral parameters [1]. The authors conclude, that this reduction is comparable to the reduction of vowels and predict a clear consequence for the intelligibility of these items. Additionally, intervocalic consonants from reading material show shorter durations and reduced spectral balance in unstressed compared to stressed syllables and for a later position in the word [2]. In [2], spectral reduction is not seen as induced by speaking rate (SR), but both effects are assumed to be a common consequence of higher level information like stress.

This study examines whether spectral reduction occurs with higher SR *within* conditions of such linguistic information. As correlate for spectral reduction for voiceless fricatives, the center of gravity (CoG) is analyzed, because a decrease of constriction and air stream velocity in the vocal tract leads to a decrease in turbulent friction, so the CoG shifts towards lower values.

2. Material

Recordings from 32 German speakers (14 women, 18 men), from the Kiel Corpus of Spontaneous Speech are analyzed. The CoG is computed from the power spectrum for the duration of 30%–70% of each item (0–8 kHz). Rate information is captured by the Perceived Local Speaking Rate (PLSR) [3]. This is a linear combination of Hann windowed (620 ms) reciprocal syllable and phone durations that highly correlates ($r = 0.9$) with user judgments of perceived tempo. Only for content words stress is annotated. Items with secondary stress, directly neighboring pauses or hesitations, and realizations of /h/ are excluded. The CoG is normalized with each speaker's mean value of each condition (allophone, stress, function/content word). Separated ANCOVAs are made for every such condition. Control variables are gender and position in the syllable.

3. Results

For all examined places of articulation, there are significant lower CoGs for increased PLSR (about 100 Hz between very slow and very fast items). Non-significant results are found only for /f/ in unstressed and function word condition (see Table 1). Results with phone durations are not listed here, because the main difference is a non-significant result for /f/. The position in the syllable is not relevant in any condition. Gender is only significant for stressed /ç/ and /χ/, where the significant lowering in CoG for increasing rate is even stronger for women.

Table 1: Rate dependent CoG; only significant results. Estimates [Hz] are for 1 PLSR (transformed to about 1 $Syllable/sec.$).

allophone	condition	estimate	F-value	DF
f	stressed	-15, 40	4.92*	489
f	stressed	-23, 45	54.95***	2248
	unstressed	-35, 88	29.82***	564
	function w.	-14, 22	10.56**	505
s	stressed	-4, 29	8.11**	2596
	unstressed	-26, 33	79.82***	2690
	function w.	-17, 91	77.16***	3374
ç	stressed	-43, 51	79.99***	708
	unstressed	-35, 88	10.43**	704
	function w.	-14, 22	91.96**	1615
x	stressed	-42, 29	6.46*	40
	unstressed	-17, 28	16.88***	166
	function w.	-51, 12	16, 28***	99
χ	stressed	-26, 04	23.58***	315
	unstressed	-52, 89	48.98***	870
	function w.	-42, 25	115.73***	607

4. Discussion and Conclusion

For increasing local speaking rate, within-speaker spectral reduction is confirmed within conditions of stress and part of speech. However, as the observed reduction is small compared to differences between individual speakers and places of articulation, perceptual relevance seems doubtful. It was not studied here, if spectral reduction is a consequence of rate, or if the correlation found has its origin in further linguistic information.

5. References

- [1] R. J. J. H. van Son and L. C. W. Pols, "An acoustic description of consonant reduction," *Speech Communication*, vol. 28, pp. 125–140, 1999.
- [2] R. J. J. H. van Son and J. P. H. van Santen, "Duration and spectral balance of intervocalic consonants," *Speech Communication*, vol. 47, pp. 100–123, 2005.
- [3] H. R. Pfützinger, "Local speech rate perception in German speech," in *Proc. of the 14th ICPhS, San Francisco*, 1999, pp. 893–896.