

Speaker and dialect effects in the dynamics of speech temporal characteristics

Volker Dellwo & Dario Brander

Phonetics Laboratory, Department of Comparative Phonetics, University of Zurich

volker.dellwo@uzh.ch

It has been repeatedly demonstrated that speakers vary in their speech rhythmic characteristics and that such characteristics might be cues to the identity of a speaker and as such relevant to forensic speaker identification (Leeman et al., 2014, Dellwo et al., 2012). A shortfall with measures of speech rhythm so far is that they are based on a durational characterization of speech intervals (e.g. a syllable, a vocalic or a consonantal interval) that is averaged over the entire utterance. For example, typical measures of speech rhythm are based on standard deviations of speech intervals (e.g. the standard deviation of consonantal intervals, Ramus et al., 1999) or the average differences between syllables in a phrase (e.g. the Pairwise Variability Index, Grabe and Low, 2002). This does not take into account the dynamics with which speakers might vary temporal characteristics of speech over the course of an utterance.

To test whether there is reason to believe that inter-speaker variability exists in the dynamics of syllable durations within an utterance we analyzed the syllable durations in 256 sentences produced by 4 male speakers of Swiss German (64 sentences each) from two dialect regions (2 Bern, 2 Zurich). The sentences were a Swiss version of the Coordinate Response Measure Corpus (Moor, 1981, Bolia, 2000) recorded in our lab in Zurich, which means that all speakers uttered structurally identical sentences of the exact same number of syllables (16) that only varied in the choice of some lexical items. To calculate the syllable duration dynamics between speakers we first calculated a proportional duration for each syllable (duration of a syllable in percent re the total duration of the utterance) and then calculated the difference in duration between consecutive syllable pairs (15 pairs); henceforth: 'Proportional syllable differences (in %)'. Figure 1 contains the mean of the proportional differences for each speaker (red = Bern, blue = Zurich; values averaged over 64 productions per speaker). A value around 0 indicates that the syllable pair was produced with about equal duration for each syllable, a positive value indicates that the first syllable in a pair was longer than the second, a negative value that the first syllable was shorter than the second.

Results revealed: (a) The largest differences were obtainable in the first part of the phrase up to syllable pair 13. This means that the phrase final part (i.e. phrase final lengthening) did not vary between speakers nor between dialects. (b) There were possible speaker and dialect effects in different parts of the sentence: Between pair 1 and 9 the differences varied strongly between speakers irrespective of their dialect. Between pair 10 and 13 the differences showed some similarities as a function of dialect.

One of the main shortfalls of this study is that it relies on highly controlled material (speakers uttered sentences of the exact same structure) and that it is based on syllable durations, a rather ambiguous durational interval in speech outside the laboratory. We are now working on methods to compare speaker specific aspects of temporal dynamics between sentences of a different structure and using different temporal intervals.

References

Dellwo, V., Leemann, A. und Kolly, M.J. (2012) Speaker idiosyncratic rhythmic features in the speech signal. In: Electronic Proceedings of Interspeech, Portland/Oregon/USA.

E. Grabe, E.L. Low (2002) Durational variability in speech and the Rhythm Class Hypothesis C. Gussenhoven, N. Warner (Eds.), Laboratory Phonology, vol. 7 Mouton de Gruyter, Berlin/New York, 515- 545.

Leemann, A., Kolly, M.-J., Dellwo, V. (2014) Speaker-individuality in supra-segmental temporal features: Implications for forensic voice comparison. In: Forensic Science International (238), 59-67.

Moore, T. J. (1981) Voice communication jamming research. In: AGARD, Conference Proceedings 311: Aural Communication in Aviation (AGARD), Neuilly-Sur-Seine, France, pp. 2:1-2:6.

F. Ramus, M. Nespors, J. Mehler (1999) Correlates of linguistic rhythm in the speech signal. In: Cognition (73), 265-292.

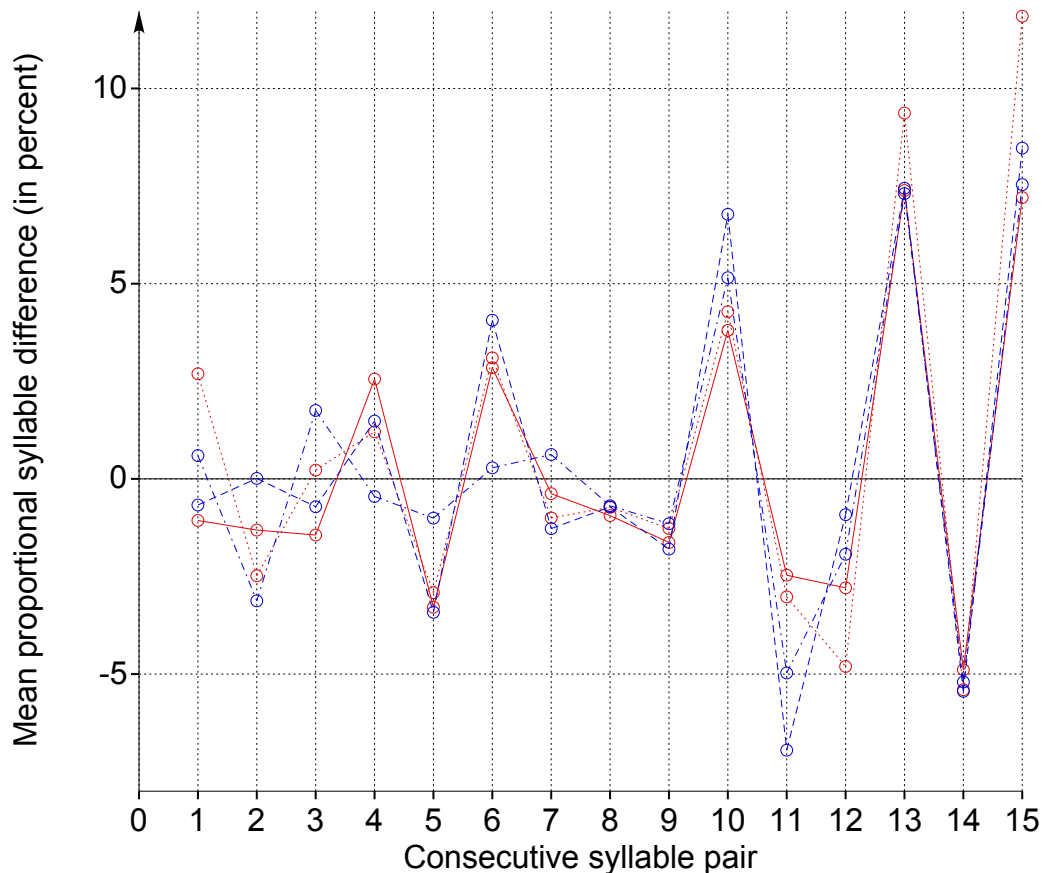


Figure 1: Graph showing the mean proportional syllable difference (in %) for each consecutive syllable pair from the first (1) to the last syllable pair (15) of the 16 syllable sentences.