Testing the effect of dialect imitation on suprasegmental temporal features

Adrian Leemann, Marie-José Kolly, Volker Dellwo
Department of Comparative Linguistics, University of Zurich
`{adrian.leemann|marie-jose.kolly}@pholab.uzh.ch,`
`volker.dellwo@uzh.ch`

Voice disguise is the intentional act of changing one's voice for the purposes of falsifying identity. The German Federal Police Office (BKA) reports that more than half of forensic cases feature a form of voice disguise (Masthoff 1996). Present research distinguishes between two types of voice disguise: electronic and non-electronic voice disguise (Künzel, 2000, Masthoff, 1996, Perro et al., 2007). Previous studies have reported that voice disguise can lead to high within-speaker variability, affecting various acoustic parameters: Hollien (1977), for instance, reported high within-speaker variability in long-term spectra under various voice disguise conditions (disguise of the speaker's own choice). Eriksson & Wretling (1997) as well as Endres et al. (1971) found that imitating another speaker is particularly achieved by adopting f0 and formant frequencies of the target speaker. High within-speaker variability makes it difficult for forensic practitioners to draw conclusions about the speakers' identity.

The objective of this contribution is to present first results of a study that tests the effect of dialect imitation as a form of non-electronic voice disguise on suprasegmental temporal features. We focus on suprasegmental temporal features because previous research in forensic phonetics has shown that these features seem relatively robust towards different variability conditions: the amount of voiced portions in the speech signal, for instance, is affected only very little by articulatory obstructions (Leemann, Hove, Kolly, Dellwo, 2014), and the amount of vocalic portions in the signal seem to remain relatively speaker-specific even if a speaker imitates a different dialect (Dellwo et al. 2009). Leemann et al. (2014) reported that a number of suprasegmental temporal features demonstrate little within-speaker variability across different speaking styles and signal distortions.

Methodologically, we proceeded as follows: 10 speakers of Zurich German were recorded at the University of Zurich. Speakers were students and showed little to no regional accent variability in Zurich German. The data was recorded in a sound-treated booth. Each speaker read 72 sentences from the TEVOID corpus (Dellwo et al., 2012, Leemann et al., 2014). Sentences typically included 15–20 syllables and were written in Zurich German. These sentences were read by the subjects and recorded as a control. The same sentences were transliterated to Bern German and read by the same subjects for the Bern dialect imitation condition. We selected Bern German since previous research has reported differences in the suprasegmental temporal features for these two dialects (Leemann, 2012, Leemann et al., 2012). We applied the following automatic measures on the labeled corpora: (1) measures that are based on intervals between amplitude peak points of a low frequency amplitude envelope (<10Hz) and (2) measures that are derived from the amount of speech voicing in the signal. In the present contribution we will present first results and discuss these findings against the backdrop of forensic phonetics.

References

Dellwo, V., A. Leemann and M.-J. Kolly. (2012). Speaker idiosyncratic rhythmic features in the speech signal. *Proceedings of Interspeech 2012*, Portland (OR).

Dellwo, V., S. Ramyead and J. Dankovicova. (2009). The influence of voice disguise on temporal characteristics of speech. Abstract presented at the IAFPA conference 2009, 02.–05.08.2009, Cambridge, UK.

Endres, W., W. Bambach and G. Flösser. (1971). Voice spectrograms as a function of age, voice disguise and voice imitation. *Journal of the Acoustical Society of America*, **49(6)**, 1842–1848.

Eriksson, A. and P. Wretling. (1997). How flexible is the human voice? – A case study of mimicry, *Eurospeech*, **97(2)**, 1043–1046.

Hollien, H. and W. Majewski. (1977). Speaker identification by long-term spectra under normal and distorted speech conditions. *Journal of the Acoustical Society of America*, **62**, 975–980.

Künzel, H. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics*, **7(2)**, 150–179.

Leemann, A. (2012). *Swiss German Intonation Patterns*. Amsterdam: Benjamins.

Leemann, A., V. Dellwo, M.-J. Kolly and S. Schmid. (2012). Rhythmic variability in Swiss German dialects. *Proceedings of Speech Prosody*, Shanghai.

Leemann, A., I. Hove, M.-J. Kolly and V. Dellwo (submitted). Testing the effect of voice disguise on suprasegmental temporal features: Corpus creation and preliminary results, Interspeech 2014.

Leemann, A., M.-J. Kolly and V. Dellwo. (2014). Speaker-individuality in the time domain: implications for forensic voice comparison. *Forensic Science International* **238**, 59–67.

Masthoff, H. (1996). A report on a voice disguise experiment. *Forensic Linguistics*, **3(1)**, 160–167.

Perrot, P., G. Aversano and G. Chollet. (2007). Voice disguise and automatic detection: review and perspectives. In Y. Stylianou et al. (Eds.), *Progress in nonlinear speech processing 2005, LNCS 4391*,101–117.