Exploring long-term formants in bilingual speakers

Willlemijn Heeren[1], *David van der Vloed*[2], *and Jos Vermeulen*[2]
[1]*Leiden University Centre for Linguistics, Leiden University, The Netherlands*
w.f.l.heeren@hum.leidenuniv.nl
[2]*Netherlands Forensic Institute, The Netherlands*
{j.vermeulen,d.van.der.vloed}@nfi.minvenj.nl

Introduction

Long-term formants (LTFs) have been forwarded as a useful feature in forensic speaker comparison (e.g. Nolan and Grigoras, 2005; Gold e.a., 2013). LTFs are assumed to be independent of individual speech sounds (Nolan and Grigoras, 2005), and earlier data support the conclusion that LTFs are language-independent (Jessen, 2010). The latter author called for more research to validate this claim, which is underlined by the finding that different speaking styles do affect LTFs (Moos, 2010). We explored if the language a bilingual is speaking influences LTFs, using forensic intercepted telephone speech (Van der Vloed et al., 2014).

Method

Recordings from twelve, male bilingual speakers of Dutch and Turkish were selected from the NFI-FRITS database (Van der Vloed et al., 2014). Recordings were pre-processed using Praat (Boersma and Weenink, 2013) to create 10-second wave files (per language and per speaker) that only included vocalic parts, following the procedure in Moos (2010). Wave files were longer than the six seconds proposed by Moos (2010) given the nature of our database: background noise and low quality may interfere with formant estimations. The first through third formant values were extracted using WaveSurfer 1.8.8p4. Only LTF2 and LTF3 were analyzed, as LTF1 is too close to the lower cutoff frequency of the phone bandwidth (see Byrne and Foulkes, 2004).

Results and discussion

To investigate if LTFs differed within speakers, but between languages, paired samples t-tests were run on both the LTF2 and LTF3 means and standard deviations. Mean LTFs within speaker and between languages did not significantly differ. Across speakers, mean LTF2 was 1418.3 Hz for Dutch and 1417.6 Hz for Turkish, and mean LTF3 was 2449.9 Hz for Dutch and 2453.8 Hz for Turkish. Standard deviations showed a difference for LTF2 ($t(11) = 2.35$, $p = .039$), but not for LTF3.

As a first comparison of the LTF distributions (LTFDs), formant histograms were compared using the Kolmogorov-Smirnov (KS) distance, either between languages within a speaker, or between speakers within a language. The KS distance is the largest absolute difference between two cumulative sample distributions. This descriptive analysis gave smaller distances for the within-speaker, between-language comparisons ($N = 12$) than for the between-speaker, within-language comparisons ($N = 66$ per language). According to Mann-Whitney-U tests, within-speaker distances for LTFD2 were smaller than between-speaker distances ($Z = -3.4$, $p = .001$), and a trend in the same direction was found for LTFD3 ($Z = -1.9$, $p = .063$).

Results are in line with previous claims that LTFs are comparable between languages, when spoken by the same speaker, and differences seem to be larger when comparing between speakers. This ties in with studies showing that LTFs may be useful in forensic speaker comparison. We aim to discuss our experiences of working with forensic speech materials, and investigate if the 10 s samples were sufficiently long for LTF estimation on such data.

References

Boersma, P. and Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.42, retrieved 2 March 2013 from http://www.praat.org/.

Byrne, C. and Foulkes, P. (2004). The 'mobile phone effect' on vowel formants. *Journal of Speech, Language and the Law* **11***,* 83-102.

Gold, E., French, P. and Harrison, P. (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. *Proceedings of Meetings on Acoustics*, Vol. 19.

Jessen, M. (2010). *Workshop Langzeitformantenanalyse*. BKA, Wiesbaden, 28 April 2010.

Moos, A. (2010). Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. *The Phonetician* **101***,* 7–24.

Nolan, F. and Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *Journal of Speech, Language and the Law* **12**, 143–173.

Van der Vloed, D. L., Bouten, J. S. and Van Leeuwen, D.A. (2014). NFI-FRITS: A forensic speaker recognition database and some first experiments *Proceedings of Odyssey Speaker and Language Recognition Workshop 2014*, Joensuu, Finland, June 16-19, 2014, pp. 6-13.