

## On the effect of changes in fundamental frequency and vocal effort in naive listener's speaker recognition

Anja Geumann

Department of Linguistics, Goethe-University, Frankfurt, Germany  
geumann@em.uni-frankfurt.de

One of the phonetic parameters quoted as especially relevant in speaker recognition is mean fundamental frequency ( $f_0$ ). As was described by Foulkes and Barron (2000) naive listeners find it easier to recognize speakers whose  $f_0$  is noticeably higher or lower than the average  $f_0$ . Additionally, a high standard deviation or high coefficient of variation of  $f_0$  can be a good cue in speaker recognition. A well observed fact in forensic phonetics is the effect of vocal effort changes on the height of  $f_0$  (see e.g. French 1998, Jessen, Köster, Gfroerer 2005). Based on these observations a speaker recognition task with speech produced at different levels of vocal effort one could expect that naive listeners will perform in one of the following ways

Listeners have more difficulties in matching speaker speaking at different levels of vocal effort if differences in mean  $f_0$  are high, because this important cue is changed.

Listeners recognize speakers better that have large differences in mean  $f_0$  between different levels of vocal effort, because they show high deviations from standard mean  $f_0$ , interpreted as high standard deviation across vocal effort settings.

An experiment was conducted to test these two contradicting assumptions.

Data were taken from the Pool 2010 corpus (see for a detailed description Jessen, Köster, Gfroerer 2005) The material consisted of studio quality, read speech, at two levels of vocal effort (neutral and Lombard setting) The higher vocal effort was induced by presenting 80 dB noise over headphones.

In the neutral setting all male speakers had a mean  $f_0$  between 100-130 Hz. All of them showed clear differences in vocal effort between the two settings, defined by mean overall intensity. One subgroup was defined by a fairly small change in  $f_0$  (<5-20 Hz higher), the other by much higher  $f_0$  in the Lombard setting (40-60 Hz higher).

Speech samples consisted of a short portion of identical read speech and the listeners were asked to match speakers from the neutral to the Lombard setting. The test design was open, foils were used. All analyses were performed using Praat (Boersma and Weenink 2009). The results will be discussed with respect to the relevance of  $f_0$  for speaker recognition.

### References

- Boersma, Paul and Weenink, David (2009). Praat: doing phonetics by computer (Version 5.1.04) [Computer program]. Retrieved April 4, 2009, from <http://www.praat.org/>
- Jessen, M., O. Köster and S. Gfroerer (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, vol. 12, 174-213.
- Foulkes, P. and A. Barron (2000) Telephone speaker recognition amongst members of a close social network. *International Journal of Speech, Language and the Law*, vol. 7, 180-198.
- French, P. (1998) Mr Akbar's nearest ear versus the Lombard reflex: a case study in forensic phonetics. *Forensic Linguistics* 5(1), 58-68.