Using the smartphone application 'Voice Äpp' to collect speech population data: implications for forensic phonetics

Ingrid Hove[1], *Adrian Leemann*[1], *Marie-José Kolly*[1], *Volker Dellwo*[1], *Jean-Philippe Goldman*[2], *Ibrahim Almajai*[2], *Daniel Wanitsch*[3]

[1] *Phonetics Laboratory, Department of Comparative Linguistics, University of Zurich, Switzerland*

ingrid.hove@uzh.ch, {adrian.leemann|marie-jose.kolly}@pholab.uzh.ch, volker.dellwo@uzh.ch

[2]*Department of Language Sciences, University of Geneva, Switzerland*

{jean-philippe.goldman|ibrahim.almajai}@unige.ch

[3]*iBros.ch LTD, Aarau, Switzerland*

dani@ibros.ch

The smartphone application *Voice Äpp*, which is currently in development, aims at providing its users with scientifically sound phonetic and dialectological information on their dialect and their voice and on general aspects of speech. For forensics, the users' recordings provide a valuable database for extracting phonetic population data.

- The application's "dialect profile" functionality is designed to determine users' dialect based on their pronunciation of 15 words using automatic speech recognition (cf. Kolly & Leemann, accepted). Since in this functionality the algorithm for dialect recognition is based on Swiss German data, this part can only be used by German speaking Swiss. The other two functionalities work for all users who understand German.
- The aim of the application's "voice profile" functionality is that users get to know characteristics of their own voice. After having recorded a given sentence in their dialect, users are shown histograms displaying their fundamental frequency and articulation rate in comparison to all of the previous users of the application.
- In the application's "infotainment" functionality the user can learn about different aspects of speech in a playful way, for example by listening to different kinds of hearing impairments or by experiencing the "McGurk effect" (McGurk & MacDonald, 1976) and the "cocktail party effect" (Handel, 1989).

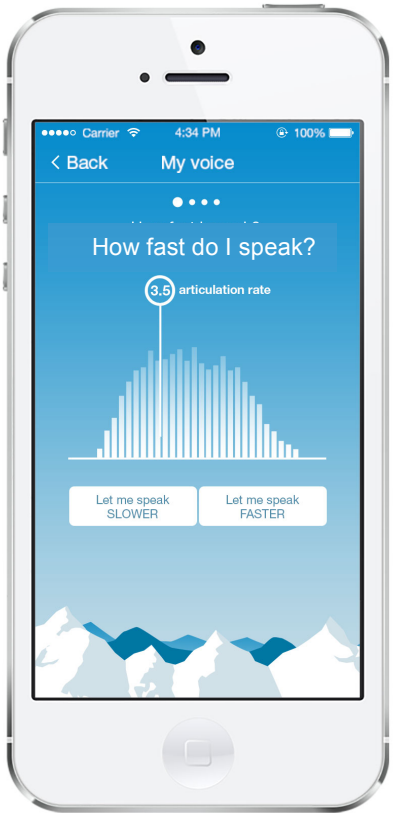

Figure 1 Screen shot showing the user's articulation rate

From a scientific point of view, *Voice Äpp* allows crowdsourcing of population data which has important implications for forensic voice comparison research. Acoustic analyses of the users' recordings will allow unprecedented insights on the areal distribution of speech signal parameters such as fundamental and formant frequencies, temporal characteristics of segments, and speaking rate. Forensic phonetic research requires population data from a large set of speakers. Until now, population statistics only exist for certain languages (English, Standard German) and typically are based on the data of around 50–100 speakers (Künzel, Masthoff & Köster, 1995; Jessen, 2007). When collecting data through crowdsourcing, certain parameters are not controllable. This disadvantage is compensated by the large amount of data we are expecting based on our experience with the predecessor application *Dialäkt Äpp* (Kolly & Leemann, accepted).

References

Boersma, P. and D. Weenink. Praat: Doing phonetics with computers, www.praat.org, accessed 31 Mar 2014.

Kolly, M.-J. & Leemann, A. (accepted). *Dialäkt Äpp*: communicating dialectology to the public – crowdsourcing dialects from the public. In: Leemann, A., Kolly, M.-J., Schmid, S. & Dellwo, V. (Eds.). *Trends in Phonetics in German-speaking Europe*, Bern/Frankfurt: Peter Lang.

Künzel, H., Masthoff, H., & Köster, J. (1995). The relation between speech tempo, loudness, and fundamental frequency: an important issue in forensic speaker recognition. *Science and Justice*, *35*, 291–295.

Jessen, M. (2007). Forensic reference data on articulation rate in German. *Science and Justice,* 47, 50–67

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 746-748.

Handel, S. (1989): *Listening. An Introduction to the perception of auditory events*. Cambridge, Mass.: MIT Press.