



Call center database creation for the needs of speaker and emotion recognition

Magdalena M. Jastrzębska, Szymon Liberski, Maksymilian Piechanowski
(Adam Mickiewicz University, Poznań)

In this paper the project for the purpose of building speech recognition and emotion recognition systems for Polish language is presented.

The core of the project constitutes a database of phone calls recordings. These are spontaneous speech samples in a form of dialogs between people calling the alarm phone number and police officers. Target scope of the database will exceed one hundred voices. The selected recordings have been annotated in respect to the verbal, quasi-verbal and nonverbal sounds, background noises, syntax, time, dialog functions, intonation and emotional coloring. For the step of annotation a freeware tool Transcriber has been used.

The preliminary specification for the annotation has been tested on a sample of selected recordings. Some improvements have been executed. Two of them seem the most salient. Replacing three with five step classification when annotating emotion dimensions as well as specifying condition for setting breakpoints especially those marking overlapping speech.

Future work will include the analysis of the distribution of fundamental frequency: F0 average, range and variability for groups of speakers divided according to sex and emotional state. For needs of speaker recognition, aside from aggregated F0 measurements, such characteristics as phones, diphthongs and nasals formant frequencies will be covered, as well as VOT and breathes distribution. All the data gathered from F0 analysis and speaker identification will be supplemented with results from perceptual annotation of emotions using three-dimensional approach applied with five step scale for each of the dimensions.

The tools used for the analysis part of the project will include: Statistica, Praat, Wavesurfer. Target scope of the database will exceed one hundred voices. As the research is being carried out at present the results will appear and will be attached shortly.

The project is supported by The Polish Scientific Committee (Project ID: R00 035 02), Technologies for processing and distributing verbal information in internal security systems, run by FUAM and the Institute of Linguistics UAM , Department of Phonetics.

References:

- Braun, A – Köster J.P. (Eds.) 1995. *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag. (Beiträge zur Phonetik und Linguistik, 64), pp. 49-55.
- Esling, J.H. 1994. *Voice quality*. In: R.E. Asher & J.M.Y. Simpson (eds). *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press.
- Hollien, H. 2001. *Forensic Voice Identification*. Academic Press.
- Jessen, M. – O. Koster – S. Gfroerer. 2005. *Influence of vocal effort on average and variability of fundamental frequency*. *Speech, Language and the Law*, vol. 12, no. 2, pp. 174–213.

- Nearey, T.M. – P.F. Assmann – J.M. Hillenbrand. 2002. *Evaluation of a Strategy for Automatic Formant Tracking*. J. Acoustical Soc. A, er., Vol. 112.
- Rose, P. 2002. *Forensic Speaker Identification*. London: Taylor & Francis.
- Scherer, K. 2005. *What are emotions? And how can they be measured?* Social Science Information, 44(4), 695-729.
- Scherer, K.R. – E. Dan – Flykt, A. 2006. *What determines a feeling's position in three-dimensional affect space? A case for appraisal*. Cognition and Emotion, 20, 92–113.
- Trautmüller, H. – A. Eriksson. 1993. *The frequency range of the voice fundamental in the speech of male and female adults*. Unpublished manuscript, Stockholm University. Retrieved from <http://www.ling.su.se/staff/hartmut/aktupub.htm>
- Ververidis, D. – C. Kotropoulos. 2006. *Emotional speech recognition: Resources, features, and methods*. Speech Communication 48(9) 1162–1181.