Advancing corpus-based analyses of spontaneous speech: Switch to GECO!

Antje Schweitzer, Natalie Lewandowski & Grzegorz Dogil

Institute for Natural Language Processing, Stuttgart University

`<firstname>.<lastname>@ims.uni-stuttgart.de`

We present the GECO (**GE**rman **CO**nversations) database, which was designed for investigating phonetic convergence in German and its correlation with personality traits and scores of mutual likeability and esteem. The data can more generally be used to investigate aspects of interlocutors' behavior in free, spontaneous dialogs, including social aspects. The database will be freely available for non-commercial use and to our knowledge will be by far the largest German database of this type. GECO consists of 46 dialogs of approx. 25 minutes length each, between previously unacquainted female subjects. It further contains results of personality tests aimed at subjects' extraversion, acting behavior, other-directedness and sensitivity to social cues (all collected once per subject), as well as participants' mutual ratings in terms of competence and likeability (collected after every dialog). 22 dialogs took place in a unimodal (UM) setting, were participants were separated by a solid wall and could not see each other, while the remaining 24 dialogs were recorded with subjects facing each other (multimodal setting, MM). All dialogs were recorded in high quality (separate channels, 16 bit@48 kHz) in a sound-attenuated booth using AKG HSC271 headsets with rubberfoam windshields. In the MM setting, a transparent screen ensured sufficient speaker separation between the two channels. We had 12 speakers in the UM condition, and 8 (7 from UM, 1 additional speaker) in the MM condition, which was recorded 5 months later. All subjects were females between 20 and 30 years of age, mostly students. They were paid for each dialog they participated in. Subjects were naïve to the research questions; they were told that the purpose of the study was to research how small talk between strangers works. They were provided with a list of potential topics to ease conversation, but were explicitly told that they were completely free to choose other topics as well.

Most recent studies on phonetic convergence and imitation use rather controlled and limited speech material, often without real conversational interaction, or focus on only specific target words or phrases in conversations (Delvaux & Soquet 2007, Babel 2010, Abrego-Collier et al. 2011, Kim et al. 2011, Nielsen 2011, Babel 2012, Pardo et al. 2012). Few recent studies use larger-scale fully annotated corpora such as the quasi-spontaneous Columbia Games Corpus (Levitan & Hirschberg 2011), which does not provide social or personality factors. However, social factors are assumed to be central in convergence. The GECO database was conceived to close this gap, providing social data in addition to large-scale fully annotated recordings of high audio quality as a basis for corpus-based approaches to phonetic convergence or other aspects of conversational behavior.

The GECO database was automatically processed using a number of tools: Starting from manual annotations, which included hesitations, filled pauses, and restarts, we automatically generated expected pronunciations using the Festival speech synthesis system (www.festvox.org) including German CELEX (Baayen et al. 1995) and an in-house morphology component to alleviate the high number of out-of-vocabulary words, which is due to the high morphological productivity of German. The spontaneous nature of our data called for the prediction of pronunciation variants in addition to the canonical pronunciations generated by our system. We then annotated all data on the segment, syllable and word level using forced alignment (Rapp, 1995), letting the alignment tool decide where variants were used instead of canonical forms. We then parameterized the F0 contours and automatically generated prosodic annotations (German GToBI labels) using classifiers trained on read data. Preliminary results indicate good precision but low recall. The quality of the automatically generated labels, which is of course not comparable to that of manual annotations, is subjectively good enough to be valuable as additional information to complement continuous prosodic parameters for instance. The resulting corpus amounts to 20.7 hrs. of dialog (two channels), with approx. 250,000 words, 360,000 syllables, 870,000 phones, 46,000 pitch accents, and 28,000 phrase boundaries. For illustration we will present experimental results (convergence of vowel formants, articulation rate, and backchannels) based on this corpus.

References

C. Abrego-Collier, J. Grove, M. Sonderegger, & A.C. L. Yu, 2011. Effects of speaker evaluation on phonetic convergence. *Proc. 17th ICPhS Hong Kong*, 192-195. / H. Baayen, R. Piepenbrock, & L. Gulikers, 1995. *The CELEX lexical database - Release 2*, Centre for Lexical Information, Max Planck Institute for Psycholinguistics, Nijmegen. / M. Babel, 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39(4), 437-456. / M. Babel, 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation, *J. Phon.* 40(1), 177-189. / V. Delvaux & A. Soquet, 2007. The influence of ambient speech on adult speech productions through unintentional imitation, *Phonetica* 64, 145-173. / M. Kim, W. S. Horton, & A. R. Bradlow, 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *J. Lab. Phon* 2, 125-156. / R. Levitan & J. Hirschberg, 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proc. 12th Interspeech Florence*, 3081-3084. / K. Nielsen, 2011. Specificity and abstractness of VOT imitation. *J. Phon.* 39(2), 132-142. / J. S. Pardo, R. Gibbons, A. Suppes, & R. M. Krauss, 2012. Phonetic convergence in college roommates. *J. Phon.* 40(1), 190-197. / S. Rapp, 1995. Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models - An aligner for German. In: *Proceedings of ELSNET Goes East and IMACS Workshop "Integration of Language and Speech in Academia and Industry",* (Moscow, Russia).