Spontaneous speech database for investigating phonetic convergence

Antje SCHWEITZER (University of Stuttgart)
Natalie LEWANDOWSKI (University of Stuttgart)
Grzegorz DOGIL (University of Stuttgart)

We describe a speech database which is being built in a study on phonetic convergence in spontaneous German conversations, with a main focus on prosody. In order to gather spontaneous speech data, with naturally occurring convergence (as opposed to controlled "imitation" in laboratory conditions), 24 dialogs with 12 different female speakers were recorded.

Many recent studies on convergence and imitation use rather controlled and limited speech material as a basis for investigating convergence. For instance, several studies focus on single word or sentence repetition (Nielsen 2007, Nielsen 2008, Babel 2009, Abrego-Collier et al. 2011). Delvaux & Soquet (2007) examine no conversational interaction but the influence of ambient speech. Pardo (2006) measures the perceptual similarity of a limited number of target words. Few current studies use quasi-spontaneous corpora such as the Columbia Games Corpus (Levitan & Hirschberg 2011) or data from interaction in a picture task (Kim et al. 2011), or fully spontaneous speech (De Looze et al. 2011). However, none of their corpora provide social factors. Gravano et al. (2011) try to overcome this lack by collecting ratings for dialog and social behavior in the Columbia Games Corpus from independent listeners in a study on turn-taking behavior which does not address convergence; however, it is not possible to collect the interactants' own mutual ratings so long afterwards.

Social behavior and attitudes are of great importance in convergence research since it is well accepted that the degree of convergence is related to social factors (Giles&Smith 1979, Giles et al. 1991, Giles & Powesland 1997, Shepard et al. 2001, Giles & Ogay 2006). For instance, Natale (1975) in one of the earliest studies on convergence correlated speakers' behavior with their scores on a "social desirability scale". Likewise, Street (1984) for instance correlated the degree of convergence with speakers' mutual ratings of social attractiveness and competence. To cater for such social factors in the present database, each speaker completed a questionnaire testing her social desirability score according to a German short version of the social desirability scale as well as their score on a scale of self-monitoring (Collani & Stürmer 2009). These tests were administered once for each speaker. In addition, after each conversation, speakers completed another questionnaire rating their assessment of the dialog (in terms of ease, relaxedness, effort, etc.), of their conversational partner (in terms of likeability, competence, etc.), and of their own attitude (in terms of feeling well, feeling superior/inferior to their interlocutor, etc.).

The present corpus will consist of 24 female-female dyads of approx. 25 minutes length each. Each speaker talked to 4 different interlocutors. No pair of interactants knew each other beforehand. Speakers where 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. The dialog pairings were designed to create a network in which each conversation constitutes a link between the involved speakers. The network was set up in a way that speakers shared as many interlocutors as possible without creating groups with more links than others. Thus, no two speakers shared the same 4 interlocutors, but speakers who talked to each other are guaranteed to have at least one interlocutor in common. This allows for comparing the behavior of each pair of interactants to their behavior in conversations with at least one other shared interlocutor. The recordings will be automatically aligned given the orthographic transcription. In the course of the project, the database will be automatically

labelled for intonation events according to GToBI(S) (Mayer 1995). As of today, 22 dialogs have been recorded, and we have started to generate the orthographic transcriptions.

First analyses of turn-taking behavior in the database were carried out using annotations which were generated completely automatically using Praat's (Boersma&Weenink, 2011) silence detection. From these, we automatically identified backchannels, assuming that each utterance of a speaker which is shorter than one second and which occurs in between utterances of the other speaker is a backchannel. We informally verified that this identified backchannels reasonably well.

We normalized the number of backchannels by interlocutor vocalization duration. We conducted a two-way analysis of variance with factors speaker and interlocutor. Here, only the effect of speaker was significant ($F(11,21)=9.53, p<0.001$), while the interlocutor effect was not significant ($F(10,21=1.43$, $p=0.23$). This indicates that the number of backchannels produced is speaker-specific and does not change significantly across conversations with different interlocutors, i.e. speakers do not seem to converge with respect to the number of backchannels produced.

Further and more detailed accounts of turn-taking behavior will be available for presentation at the conference.

References

Abrego-Collier, C., Grove, J., Sonderegger, M., Yu, A. C. L., 2011. Effects of speaker evaluation on phonetic convergence. In: Proceedings of the 17th International Congress of Phonetic Sciences, pp. 192–195.

Babel, M. E., 2009. Phonetic and social selectivity in speech accommodation. Ph.D. thesis, University of California, Berkeley.

Boersma, Paul & Weenink, David (2011). Praat: doing phonetics by computer [Computer program]. http://www.praat.org/

Collani, G., von, & Stürmer, S., 2009. Deutsche Skala zur Operationalisierung des Konstrukts Selbstüberwachung (Self-Monitoring) und seiner Facetten. In: A. Glöckner-Rist (ed.), Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 13.00. Bonn: GESIS.

De Looze, C., Oertel, C., Rauzy, S., Campbell, N., 2011. Measuring Dynamics of Mimicry by Means of Prosodic Cues in Conversational Speech. In: Proceedings of the 17th International Congress of Phonetic Sciences, pp. 1294-1297.

Delvaux, V., Soquet, A., 2007. The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. In: Phonetica 64 (2-3), pp. 145–173.
Giles, H., Smith, P. M., 1979. Accommodation theory: Optimal levels of convergence. Language and Social Psychology, 45–65.

Giles, H., Coupland, N., Coupland, J., 1991. Accommodation theory: Communication, context and consequence. In: Giles, H., Coupland, N., Coupland, J. (Eds.), Contexts of Accommodation. Cambridge University Press, pp. 1–68.

Giles, H., Powesland, P., 1997. Accommodation theory. In: Coupland, N., Jaworski, A. (Eds.), Sociolinguistics - A Reader and Coursebook. Palgrave, pp. 232–239.

Giles, H., Ogay, T., 2006. Communication accommodation theory. In: Whaley, B., Samter, W. (Eds.), Explaining communication: Contemporary theories and exemplars. Mahwah, NJ: Lawrence Erlbaum, pp. 293–310.

Gravano, A., Levitan, R., Willson, L., Benus, S., Hirschberg, J., Nenkova, A., 2011. Acoustic and Prosodic Correlates of Social Behavior. In: Proceedings of Interspeech 2011 (Florence, Italy), pp. 97-100.

Kim, M., Horton, W. S., Bradlow, A. R., 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. Laboratory Phonology. In: Laboratory Phonology 2 (1), pp. 125–156.

Levitan, R., Hirschberg, J., 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: Proceedings of Interspeech 2011 (Florence, Italy), pp. 3081-3084.

Mayer, J., 1995. Transcription of German intonation—The Stuttgart system. Technical Report, Institute of Natural Language Processing, University of Stuttgart.

Natale, M., 1975. Social desirability as related to convergence of temporal speech patterns. Perceptual and Motor Skills 40, 827–830.

Nielsen, K., 2008. Word-level and feature-level effects in phonetic imitation. Ph.D. thesis, University of California, Los Angeles.

Nielsen, Y. K., 2007. Implicit phonetic imitation is constrained by phonemic contrast. In: Trouvain, J., Barry, J. W. (Eds.), Proceedings of ICPhS 2007 (Saarbrücken, Germany). pp. 1961–1964.

Pardo, J. S., 2006. On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America 119 (4), 2382–2393.

Shepard, C. A., Giles, H., Le Poire, B. A., 2001. Communication accommodation theory. In: Robinson, W. P., Giles, H. (Eds.), The New Handbook of Language and Social Psychology. John Wiley & Sons, pp. 33–78.

Street, R., 1984. Speech convergence and speech evaluation in fact-finding interviews. Human Communication Research 11 (2), 139–169.