ATTENTION, PLEASE! EXPANDING THE GECO DATABASE

Antje Schweitzer, Natalie Lewandowski, Daniel Duran, Grzegorz Dogil

Institute of Natural Language Processing, Stuttgart University
{schweitzer,lewandowski,duran,dogil}@ims.uni-stuttgart.de

ABSTRACT

This paper describes current and future contents of the (**GE**rman **CO**nversations) conversations database and promotes investigating the role of attention in phonetics research. GECO is freely available for non-commercial use. It consists of conversations of high-audio quality between female subjects, together with results of personality tests of each participant, and participants' ratings of each other and of the conversation. To our knowledge it is currently the largest German database of this type. This corpus will be doubled in size by adding more dialogs in the next two years, and these new speech data will be complemented by results of several attention tests. Some of these tests will follow established test paradigms, but we also suggest a new, less artificial paradigm for testing attention. We describe the existing GECO corpus as well as the future additions including the proposed test in this paper.

Keywords: Speech database, Conversations, Phonetic Convergence, Attention

1. INTRODUCTION

The GECO database was originally designed for investigating phonetic convergence in German. The aim is to investigate the role of a variety of acoustic parameters in convergence, as well as to test the relevance of social and personality factors [23, 24, 25]. 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 [6, 3, 9, 1, 8, 16, 4, 18, 17]. Few recent studies use larger-scale fully annotated corpora such as the quasi-spontaneous Columbia Games Corpus [11]. This corpus however does not provide social or personality factors, which are assumed to be central in convergence. While it is possible to add such variables to an existing corpus later, by asking listeners about how they perceive these factors in the recorded conversations, as [10] did for the Columbia Games Corpus, to our knowledge no existing corpus contains self-assessed social scores.

GECO was conceived to close this gap, providing social data in addition to large-scale fully annotated recordings of completely spontaneous speech with high audio quality as a basis for corpus-based approaches to phonetic convergence. Even though GECO was targeted at investigating convergence, it is highly interesting for any kind of research on conversation behavior in general, and even more so for assessing the influence of social and personality factors on conversational speech.

We will now set out to investigate the role of attention as a new factor in convergence, and to this end, we will add more conversations from new participants to the GECO corpus. These subjects will be tested for attention aspects, and their results will be included in the new corpus, together with the corresponding social and personality data, as before.

We describe the corpus in some detail in the following section, before sketching the attention tests that subjects will participate in during the next phase in section 3.

2. THE CURRENT GECO DATABASE

2.1. Recording conditions & participants

GECO consists of 46 dialogs of approx. 25 minutes length each, between previously unacquainted female subjects. 22 dialogs took place in a unimodal (UM) setting, where 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.

There were 12 speakers in the UM condition. We recruited the same speakers again in cases where they were still available, to be able to compare speakers across conditions. Of the 12 speakers, 7 returned for the MM condition, and we recruited one additional speaker. This meant that some dialog pairings from the UM condition are repeated in the MM condition. We recorded the MM condi-

tion 5 months after the UM condition, and we assume that any convergence effects were lost by then. All dialogs are in Standard German, with dialectal coloring for some speakers of Swabian origin. 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. The resulting corpus amounts to 20.7 hrs. of dialog.

2.2. Social and personality data

GECO further contains results of a test aimed at several personality aspects (collected once per subject). Participants were tested using the scale reported in [5]. This scale is a German adaptation of Snyder's self-monitoring scale [28]. It uses German translations of the English items from the original scale as well as some new items. The new items extend the dimensions to include sensitivity to expressive behavior and social cues, so that the German version now provides results on four instead of three scales, viz. sensitivity to expressive behavior and social cues as well as acting behavior, other-directedness and extraversion.

The database further contains participants' mutual ratings in terms of competence and likeability (collected after every dialog), and their assessment of the conversation in terms of pleasantness, atmosphere, and ease of conversation. They also indicated how they felt during the conversation (self-confidence, nervousness, and superiority/inferiority towards the conversation partner).

2.3. Annotation

The GECO database was automatically processed using a number of carefully selected tools. We first manually transcribed the dialogs orthographically for each speaker in each dialog, including hesitations, filled pauses, and restarts. In the transcripts these are distinguishable from fluent speech by markers such as "…" or "–" and could be filtered if necessary. In order to facilitate further automatic analyses on the syntactic level (e.g., POS tagging, syntactic parsing, lexicon lookup), the manual transcripts obey standard German orthography where possible. Deviations from the standard were modeled on the phonetic level: We first generated canonical pronunciations using the Festival speech synthesis system (www.festvox.org) including an extensive pronunciation lexicon German CELEX [2] and an in-house morphology component to alleviate the high number of out-of-vocabulary (OOV) words. The high number of OOV words is due to the high morphological productivity of German. We implemented a component that takes canonical pronunciations as input and predicts pronunciation variants to model the reductions often seen in spontaneous as opposed to read speech, as well as the dialectal coloring observed for some speakers. We annotated all data on the segment, syllable and word level using forced alignment [19], letting the alignment tool decide where variants were used instead of canonical forms.

We parameterized the F0 contours and automatically generated prosodic annotations according to the Stuttgart GToBI system [15] using classifiers trained on read data [22]. The prosodic annotations are highly experimental, as they were trained on read news speech instead of conversational, spontaneous speech. 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 annotations then contain approx. 250,000 words, 360,000 syllables, 870,000 phones, 46,000 pitch accents, and 28,000 phrase boundaries.

2.4. Adding more conversations

The conversations to be added will be processed in the way sketched above. Our goal is to add 48 new conversations with new participants, which will approximately double the amount of speech data. Participants will again take the personality test, and rate each conversation and each conversation partner. In addition, they will participate in attention tests. We motivate the relevance of attention and elaborate the tests in more detail in the following section.

3. ATTENTION

3.1. The role of attention in convergence

[12] in her dissertation found that phonetically talented subjects in an L2 setting converged more than less talented subjects. This finding was explained by assuming an exemplar-theoretic perspective and proposing that in order to store (and retrieve) exemplars in memory along with fine phonetic detail, attention to this detail is a necessary prerequisite. An

individual's ability to pay attention to fine phonetic detail, in turn, was hypothesized to be a substrate of phonetic talent, which escapes conscious access and direct control and is located at the core of the convergence mechanism (alongside individual personality features which may influence adaptation).

This hypothesis was supported by a post-hoc analysis of the convergence results in [12] involving data from a classical mental flexibility test (Simon Test [27]), which revealed a positive correlation between the two dimensions. This test requires subjects to quickly re-tune their attention to a changing scenario and to suppress habitual answers. The better the subjects performed in the Simon task, the more phonetic convergence they displayed during the dialogs [13]. Although the classical Simon task is a non-verbal test, it is assumed to employ a large neuronal network, also overlapping with areas crucial for other attention-demanding tasks [21]. This could explain why the test seems to capture an essential dimension for speech convergence as well.

We take this as an indication of the role attentional processes play in phonetic talent and consequently also in convergence. However, the applied test was not language-based and cannot deliver any detailed insights about which type of attention or which precise attention mechanism might contribute to individual differences in phonetic convergence. We will address this aspect in more detail in the same way we have addressed the relevance of social factors in convergence, i.e. we will record more dialogs and collect data on speakers' attentional capacities, in addition to the social and personality data. We can then correlate the attentional capacities and convergence, as well as their interaction with the previously established social factors.

3.2. Testing attention

Thus a first step is to develop tests aiming at attention to phonetic detail. Segalowitz [26] proposes that two processes contribute to overall fluency in speech (in both decoding and producing), access fluidity (AF) and attention control (AC). AC is defined as the ability to focus and refocus attention on different semantic levels (local vs. global meaning relations).

While Segalowitz focuses on shifting between local and global meaning access, we propose that attention control can also be involved in switching between various dimensions of the speech signal, for instance between detailed acoustic shape and meaning. AC is usually tested in an alternating runs paradigm [20], where participants make a series of judgments in two alternating differing tasks. The other process – access fluidity – concerns the speed and/or automaticity of connecting words to their meaning. AF is usually measured by reaction times in (lexical or semantic) judgment tasks or tasks for automaticity in comprehension [26].

For the expanded database we plan to test subjects with the alternating runs paradigm and the classical Simon Task for mental flexibility. However, in addition to these well established tests, we will use a newly designed computer-based paradigm, which is described in the following section.

3.3. A computer-game experimental framework

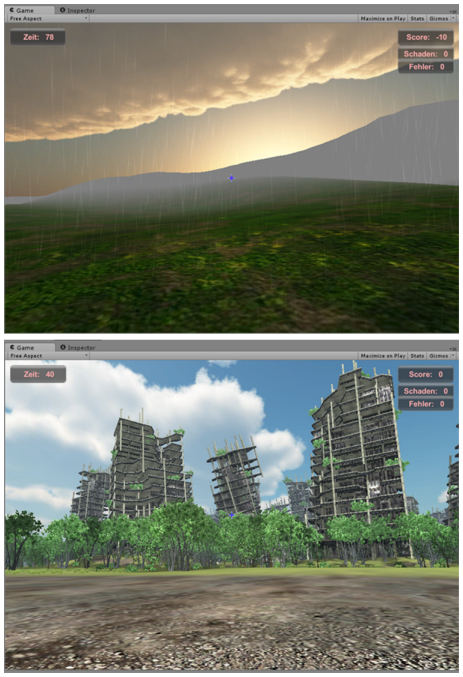
One major drawback of (e.g. lexical) judgment tasks is that they do not represent a natural scenario for attention to phonetic detail. The typical tasks used in testing AC are notoriously supervised, loaded with feedback control, and lacking in naturalness. A solution to this problem is employing a computer game scenario in which AC and AF are a natural part of the game. Computer game paradigms enjoy increasing popularity in psychological experiments [7, 14, 31]. A computer game will yield more natural data as paying attention arises as a necessity from the game scenario and requires a certain action in response to an event rather than an explicit judgment of any sort. [14] successfully used the so-called "*irf-bat*" game paradigm [30] for improving non-native speech category perception. We will use an adapted version of the irf-bat paradigm.

A first version of the adapted game has been implemented using the Unity game engine [29]. The basic principle behind the game is inspired by the irf-bat game [30]. With this state-of-the-art game engine, however, the game provides a much more immersive virtual 3D environment for the player (i.e. for the subject of the study). Screenshots of the enviroment in the adapted game are depicted in Fig. 1.

The game guides the player through three successively more difficult levels (i.e. stages or episodes of the game within a limited space and time). A science-fictional story provides a consistent background for the player, who has to navigate through a virtual environment and catch animated characters using the computer keyboard and a mouse.

As in the irf-bat study [30], there are different kinds of characters to which the player has to respond using different keys on the mouse. The characters are animated humanoid figures which are colored according to their corresponding class. Additionally, they make different sounds. However, only the significance of the different colors is pointed out to the player.

Figure 1: Screenshots of the adapted irf-bat game for testing attention. The game provides a much more immersive virtual 3D environment for the player than the original irf-bat game.

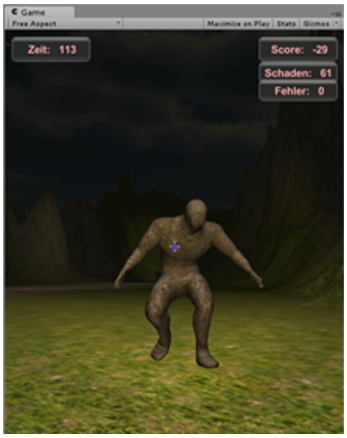


In order to score in the game, it is necessary to properly discriminate the characters, and while at the beginning of the game, they can be distinguished both visually and auditorily, in later stages, they only differ in sounds. Fig. 2 shows an example character who can not be classified by color anymore. Thus the player must figure out that it is beneficial to pay attention to the sounds. The sounds can be replaced by the experimenter to test discrimination of whatever phonetic parameters are of interest in a given study.

The three game levels are embedded within a base level (also a 3D environment to be navigated) which connects the game levels and provides a framing within the game's story. In order to assess the individual base-line performance of each player, a training level must be completed prior to the actual game levels. The training level as well as the framing base level serve an additional purpose: it has been shown that some game-internal training can reduce the a priori differences in navigation performance between experienced and inexperienced players [7].

The entire game is designed such that the total playing time is approximately 30 min or less. In our first experimental runs subjects reportedly enjoyed the game. We will present results which show that the computer game framework indeed enables

Figure 2: Screenshot of the adapted irf-bat game for testing attention showing one of the target characters which the player has to catch.



participants to recognize the relevance of the different sounds and respond to them appropriately. With these preliminary observations we show that the framework provides a suitable basis for psycholinguistic studies accompanying the empirical data from the laboratory speech recordings.

4. CONCLUSION

We propose that attention plays a role in phonetic convergence, and suggest to use an adapted version of the irf-bat paradigm [30] for testing attention in a less artificial way than in existing tests. We will extend the GECO database, doubling it in size. Participants in all new dialogs will be tested for attention using established tests as well as the adapted irf-bat test, and their results will be included in the database, which will again be freely available for non-commercial use.

5. ACKNOWLEDGEMENTS

The GECO database was created in the project *Phonetic Convergence in Spontaneous Speech* within the SFB 732 at the university of Stuttgart, funded by the German Research Foundation (DFG).

The computer game was implemented and evaluated by Lisa Lange and Bartholomäus Pfeiffer in partial fulfillment of their diploma theses at the Institute of Natural Language Processing, at the University of Stuttgart.

6. REFERENCES

[1] Abrego-Collier, C., Grove, J., Sonderegger, M., Yu, A. C. L. 2011. Effects of speaker evaluation on phonetic convergence. *Proc. ICPhS XVII, Hong Kong*.

[2] Baayen, H., Piepenbrock, R., Gulikers, L. 1995. The CELEX lexical database – release 2. Lexical Information, Max Planck Institute for Psycholinguistics, Nijmegen.

[3] Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39, 437–456.

[4] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40(1), 177 – 189.

[5] Collani, G. v., Stürmer, S. 2009. Deutsche Skala zur Operationalisierung des Konstrukts Selbstüberwachung (Self-Monitoring) und seiner Facetten. In: Glöckner-Rist, A., (ed), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. Bonn: GESIS.

[6] Delvaux, V., Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64(2-3), 145–173.

[7] Frey, A., Hartig, J., Ketzel, A., Zinkernagel, A., Moosbrugger, H. 2007. The use of virtual environments based on a modification of the computer game Quake III Arena® in psychological experimenting. *Computers in Human Behavior* 23(4), 2026–2039.

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

[9] Lelong, A., Bailly, G. 2011. Study of the phenomenon of phonetic convergence thanks to speech dominoes. In: Esposito, A., Vinciarelli, A., Vicsi, K., Pelachaud, C., Nijholt, A., (eds), *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*. Springer Berlin Heidelberg 273–286.

[10] Levitan, R. 2014. *Acoustic-Prosodic Entrainment in Human-Human and Human-Computer Dialogue*. PhD thesis Columbia University.

[11] Levitan, R., Hirschberg, J. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Interspeech Conference Proceedings* Florence, Italy. 3081–3084.

[12] Lewandowski, N. 2012. *Talent in nonnative phonetic convergence*. PhD thesis Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

[13] Lewandowski, N. 2013. Phonetic convergence and individual differences in non-native dialogs. *New Sounds* Montréal, Canada.

[14] Lim, S.-j., Holt, L. L. 2011. Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science* 35(7), 1390–1405.

[15] Mayer, J. 1995. Transcription of German intonation—the Stuttgart system. Technical report Institute of Natural Language Processing, University of Stuttgart.

[16] Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2), 132 – 142.

[17] Pardo, J. S., Cajori Jay, I., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., Krauss, R. M. 2013. The influence of role-switching on phonetic convergence in conversation. *Discourse Processes* 50, 276–300.

[18] Pardo, J. S., Gibbons, R., Suppes, A., Krauss, R. M. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40(1), 190 – 197.

[19] Rapp, S. 1995. Automatic phonemic transcription and linguistic annotation from known text with hidden markov models – an aligner for german. *Proceedings of ELSNET Goes East and IMACS Workshop 'Integration of Language and Speech in Academia and Industry'* Moscow, Russia.

[20] Rogers, R. D., Monsell, S. 1995. Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General* 124(2), 207–231.

[21] Rota, G., Reiterer, S. 2009. Cognitive aspects of pronunciation talent. *Language talent and brain activity* 1, 67–96.

[22] Schweitzer, A. 2011. *Production and Perception of Prosodic Events—Evidence from Corpus-based Experiments*. Doctoral dissertation, Universität Stuttgart.

[23] Schweitzer, A., Lewandowski, N. 2013. Convergence of articulation rate in spontaneous speech. *Proc. 14th Annual Conference of the International Speech Communication Association (Interspeech 2013, Lyon)* 525–529.

[24] Schweitzer, A., Lewandowski, N. 2014. Social factors in convergence of f1 and f2 in spontaneous speech. *Proc. 10th International Seminar on Speech Production, Cologne*.

[25] Schweitzer, A., Lewandowski, N., Dogil, G. 2014. Advancing corpus-based analyses of spontaneous speech: Switch to GECO! *Abstract at LabPhon, Tokyo*.

[26] Segalowitz, N. 2007. Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly* 41(1), 181–186.

[27] Simon, J. R., Rudell, A. P. 1967. Auditory s-r compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology* 51(3), 300 – 304.

[28] Snyder, M. 1974. Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology* 30, 526–537.

[29] Unity Technologies, 2014. Unity. Video game creation system and game engine. www.unity3d.com.

[30] Wade, T., Holt, L. L. 2005. Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America* 118(4), 2618–2633.

[31] Washburn, D. A. 2003. The games psychologists play (and the data they provide). *Behavior Research Methods, Instruments, & Computers* 35(2), 185–193.