Hindawi Publishing Corporation
EURASIP Journal on Audio, Speech, and Music Processing
Volume 2010, Article ID 835974, 2 pages
doi:10.1155/2010/835974

Editorial
Atypical Speech

Georg Stemmer,[1] Elmar Nöth,[2] and Vijay Parsa[3]

[1] *Research & Development ASR, SVOX Deutschland GmbH, Balanstrasse 73/Geb. 21, 80151 Munich, Germany*
[2] *Department of Pattern Recognition, Friedrich-Alexander University of Erlangen-Nuremberg, Martensstrasse 3, 91058 Erlangen, Germany*
[3] *National Centre for Audiology, The University of Western Ontario, 1201 Western Road, Elborn College, Canada N6G 1H1*

Correspondence should be addressed to Georg Stemmer, georg.stemmer@svox.com

Received 2 March 2010; Accepted 2 March 2010

One of the most important aspects of spoken language is its large degree of variability. Variability in speech is caused by many different sources, for instance, changes of the acoustic environment or transmission channel and differences between speakers or various speaking styles. Successful speech processing systems typically combine several different means to cope with the unwanted variability of the input signal. In the last two decades, large progress has been made in the areas of feature-normalization, speaker-independent and speaker-adaptive acoustic modeling, and robust estimation methods for statistical language models. This has led to many useful applications of speech processing, like spoken dialogue systems that are connected to the telephone network, medical dictation, broadcast news transcription, or spoken destination entry for navigation systems in the car. Unfortunately, the algorithms used in current systems for robust modeling, speaker normalization and adaptation have many limitations, in particular for speech that deviates significantly from the data in the training corpus. Atypical speakers like nonnative speakers, children, or members of the elderly population still lead to much higher error rates in state-of-the-art speech recognizers than normal, or typical, adult native speakers.

This limits the practical applications of automatic speech processing significantly. For instance, a spoken dialogue system should be able to understand any user, even if he or she belongs to the elderly population. Furthermore, the system should be able to react in an adequate manner if the user's emotional state changes. A software for computer-aided language learning needs to be able to cope with nonnative speech.

As research in the past has concentrated more on typical speech than on atypical speech, some important questions in this area are still largely unanswered. For instance, there is no good definition of the term *atypical speech* yet. The articles we present in this special issue investigate speech from speakers with disabilities, nonnative speech, children's speech, speech from the elderly, speech with emotional content and singing. For many types of variability, the reasons for the increased error rates are still unknown. Furthermore, it is unclear whether the error rates could be reduced by collecting adequate amounts of training (or adaptation) data or whether novel processing methods have to be developed. We hope that the papers in this special issue help to advance in the direction of getting an answer to these questions. The majority of the articles analyses the influence of atypical speech on automatic speech recognition performance in great detail, and different methods to reduce the error rates for atypical speech are proposed and evaluated. Two papers investigate how different voice qualities can be distinguished automatically.

Correspondingly, we have grouped the papers in this special issue into three areas.

The first area consists of papers that investigate the influence of atypical speech on automatic speech recognition performance. The article *On the impact of children's emotional speech on acoustic and language models* by S. Steidl, A. Batliner, D. Seppi, and B. Schuller investigates the influence of the emotional state of a speaker on speech recognition performance. When the speech recognizer is trained on neutral speech, the somewhat surprising result for a collection of spontaneous utterances from children is

that emphatic and angry speech, is recognized better than neutral speech. A possible explanation for this observation is that in emphatic and angry speech the words are pronounced clearly and with less intraclass variability. Therefore, they may fit on average better to the acoustic models than speech from other emotional states. In the paper *Ageing voices: the effect of changes in voice parameters on ASR performance* the authors R. Vipperla, S. Renals, and J. Frankel analyse different properties of the speech signal which may be responsible for the decrease in the accuracy of a speech recognizer for elderly speakers. Voice source parameters like jitter and shimmer change with age, but they are shown to have just a minor influence on speech recognition error rates. Instead, the authors find indications that a systematic change in the acoustic space for certain phones seems to be responsible for the decrease in speech recognition performance with increasing age of the speaker.

The second area contains articles that investigate new approaches and combinations of existing approaches to directly improve speech accuracy for atypical speech. The paper *Automatic recognition of lyrics in singing* by A. Mesaros and T. Virtanen describes the development of a system for the recognition of sung speech. Different adaptation and language model training methods are combined and lead to a speech recognition system that can be used for a query-by-singing application. *Exploring the effect of differences in the acoustic correlates of adults' and children's speech in the context of automatic speech recognition* by S. Ghai and R. Sinha has the goal to normalize different acoustic parameters to reduce the differences between speech from children and adults. The transformation of formant frequencies, speaking rate and pitch leads to significant reductions of the error rates for the young speakers when the speech recognizer has been trained on speech from adults and vice versa. The article *Optimizing automatic speech recognition for low-proficient nonnative speakers* by J. van Doremalen, C. Cucchiarini, and H. Strik describes the development of a system for computer-aided language learning. Such a system has to deal with speakers that have a very low proficiency of the foreign language, thus speech recognition error rates are very high. The authors propose to avoid automatic recognition of unconstrained input and restrict the responses that the user can give. A combination of utterance selection and verification from a list of predefined phrases is utilized in order to avoid giving confusing responses to the learner.

The third area contains articles that develop methods to analyse atypical speech and describe approaches to distinguish different voice qualities.

The article *Automatic speech recognition systems for the evaluation of voice and speech disorders in head and neck cancer* by A. Maier, T. Haderlein, F. Stelzle, E. Nöth, E. Nkenke, F. Rosanowski, A. Schützenberger, and M. Schuster describes a system that uses an automatic speech recognizer in medical rehabilitation. The intelligibility of speakers with different speech disorders is measured and quantified in an automated manner by measuring the word recognition rate of the speech recognition system. A clear correlation between the judgement of human experts and the automatically generated recognition rate is shown. The paper *Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech* by C.T. Ishi, H. Ishiguro, and N. Hagita has the goal to incorporate paralinguistic cues about certain user states (e.g., emotions) in a spoken dialogue system. The authors demonstrate that breathy and whispery voice qualities can be used to detect some of these cues. Different acoustic parameters are extracted from the speech signal to detect breathy and whispery segments in spontaneous speech.

Georg Stemmer
Elmar Nöth
Vijay Parsa