The PAVOQUE corpus as a resource for analysis and synthesis of expressive speech

Ingmar Steiner\textsuperscript{1,3}, Marc Schröder\textsuperscript{2}, and Annette Klepp\textsuperscript{2,3}

\textsuperscript{1}Cluster of Excellence “Multimodal Computing and Interaction”, Saarland University
\textsuperscript{2}Language Technology Lab, DFKI GmbH, Saarbrücken
\textsuperscript{3}Computational Linguistics & Phonetics, Saarland University
steiner@coli.uni-saarland.de

Abstract

The nature of expressive and emotional speech has garnered a mounting body of research over the past decade (Scherer, 2003; Schröder, 2009; Schuller et al., 2011, among many others); a number of research projects have been, or are being, conducted in order to investigate phonetic parameters of expressive speech and to implement their findings in technological applications. Independent scientists in phonetics and related disciplines may however share an interest in this field and the research questions it entails, open or answered but unreplicated. A significant obstacle however is the requirement for speech corpora of appropriate size and content, especially those extensively annotated with linguistic metadata; especially for German, not many such resources are available (cf. however Burkhardt et al., 2005).

This paper presents a corpus of read speech from a single male speaker of German,\textsuperscript{1} which contains five distinct speaking styles, viz. neutral, cheerful, depressed, aggressive, and a “cool, laid-back” poker style. The corpus comprises 3000 sentences, optimized for phonetic coverage; 400 of these sentences, as well as 150 domain-specific utterances, were recorded in each of the expressive styles. Phone-level segmentation is available for all of the recorded utterances, and the labels were manually checked and corrected where needed.

The corpus has been used for voice conversion (Türk and Schröder, 2010) and to create voices for expressive text-to-speech synthesis (Gebhard et al., 2008; Steiner et al., 2010), which in turn have found use in a number of studies (e.g. Scheffler et al., 2012; Székely et al., 2013). However, the data itself was never made available to the public, and so its use as a resource for the analysis of expressive speech, or as an asset for novel technological applications, was hitherto restricted. With this paper, we announce the availability of the full corpus, free of charge, under

\textsuperscript{1}Stefan Röttig, http://www.stefan-roettig.de/
a much more permissive license, in the belief that the scientific community will regard it as a valuable resource for phonetic research and other applications. In the spirit of Rosenberg (2012), we use distributed version control (Torvalds, n.d.) and peer-to-peer data mirroring (Hess, n.d.) to manage the phonetic annotations and speech data, respectively, allowing the corpus to be easily maintained and enhanced, and integrated into other projects as a submodule.

References


