MICHAEL PUCHER, NADJA KERSCHHOFER-PUHALO, DIETMAR SCHABUS, SYLVIA MOOSMÜLLER, GREGOR HOFER

FTW FORSCHUNGSZENTRUM TELEKOMMUNIKATION WIEN, ÖSTERREICH
{pucher,schabus,hofer}@ftw.at
ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN, INSTITUT FÜR SCHALLFORSCHUNG
{nadja.kerschhofer,sylvia.moosmueller}@oeaw.ac.at

Language resources for the adaptive speech synthesis of dialects

In man-machine communication, speech synthesis forms an important part. However, at present, speech synthesis systems are restricted to the synthesis of standard varieties. This restriction implies a strong limitation of possible application scenarios. Therefore, the synthesis of dialects and sociolects enables an important expansion.

Based on our previous work on the synthesis of standard Austrian German and on the synthesis of various Viennese varieties, we currently develop a speech synthesis system for two further Austrian dialects, namely the dialect of Innervillgraten (Eastern Tyrol) and the dialect of Bad Goisern (Salzkammergut, Upper Austria). As a result, a (multimodal) audio-visual synthesis of the two dialects will be provided.

The current contribution describes the steps necessary for the construction of a speech synthesis system for dialects. It will be shown how dialects for which neither a corpus nor a sufficient linguistic description exist are modelled and synthesized by means of hidden Markov models (HMMs) on the basis of a comparatively small corpus (600 utterances per dialect). In HMMs, contextualized phone sequences are modelled via in-between dependencies and via dependencies on observed acoustic features. The application of background models plus an adaptation of these models allows an automatic training of HMM models for a given speaker.

This contribution focuses on the quality of the recordings, on the methods of data collection, on the quality of the speech corpus for the synthesis, on the characteristics of the underlying “phone set”, and the necessary steps of training and of the analysis of the material.

Acknowledgment

This research was funded by the Austrian Science Fund (FWF): P22890-N23. The Competence Center FTW Forschungszentrum Telekommunikation Wien GmbH is funded within the program COMET - Competence Centers for Excellent Technologies by BMVIT, BMWA, and the City of Vienna. The COMET program is managed by the FFG.

References

[1] M. Pucher, N. Kerschhofer-Puhalo, D. Schabus, *Phone set selection for HMM-based dialect speech synthesis*. 1st Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties (DIALECTS 2011). EMNLP 2011: Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK.

[2] M. Pucher, F. Neubarth, V. Strom, S. Moosmüller, G. Hofer, C. Kranzler, G. Schuchmann, D. Schabus, *Resources for speech synthesis of Viennese varieties*. In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC), Valletta, Malta.

[3] S. Moosmüller, *Soziophonologische Variation im gegenwärtigen Wiener Deutsch*, Franz Steiner Verlag, Stuttgart (1987).