MICHAEL PUCHER, DIETMAR SCHABUS, GREGOR HOFER

FTW FORSCHUNGSZENTRUM TELEKOMMUNIKATION WIEN, ÖSTERREICH
{pucher,schabus,hofer}@ftw.at

From Viennese to Austrian German and back again – An algorithm for the realization of a variety-slider

In this talk we want to present an algorithm that allows for the automatic generation of in-between varieties, given acoustic models from multiple varieties. This is useful in speech synthesis to realize realistic voice user interfaces, since these phenomena also occur in natural language. The method can be extended to take account of switching direction and articulatory processes. We will demonstrate the method with Viennese varieties (sociolects).

With state-of-the-art speech synthesis methods it is possible to generate naturally sounding and comprehensible speech. The synthesis of varieties is challenging due to the lack of a standard orthography and the similarity of these varieties to spontaneous speech. The realization of a variety slider is a prerequisite to synthesize different aspects of varieties. Since speech synthesis will be used in many future user interfaces, it is important to have speech synthesis of many varieties. This is a general requirement for speech synthesis systems, since the existence of social and regional varieties is a global phenomenon.

The variety-slider algorithm interpolates between different acoustic models. The interpolation is performed on the level of phones, where phones are modeled by automatically learned context models. The mapping between phones of different varieties can be realized using dynamic programming and acoustic similarity metrics.

Acknowledgement

References

[1] M. Pucher, D. Schabus, J. Yamagishi, F. Neubarth, V. Strom, *Modeling and interpolation of Austrian German and Viennese dialect in HMM-based speech synthesis*. Speech Communication, Volume 52, Issue 2, February 2010, Pages 164-179.

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

[3] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, T. Kitamura, *Speaker interpolation for HMM-based speech synthesis system*, Acoust. Sci. Technol., 21 (4) (2000), pp. 199–206.