PhD proposal: Application of Speech & Sound Optimization Techniques for Real-World AAC TOBII D2, D3, D5, D6 Scenarios (WP2, WP3)

Stockholm - January 2017

Motz Benedicte

Contents

List of Figures

List of Tables

This project proposes to help people with complex communication impairment in their everyday lives. Now a days different low-tech an high-tech systems permit to assist individuals with difficulties to speak. In this proposal we focus on high-tech systems, more precisely AAC devices. Technological development permitted the high-tech AAC devices to become effective communication aids enabling someone with speech impairment to communicate their wishes or needs, by generating the user's desired speech signal. [1] shows that improvements have to be done, to enable an easier accessibility of the device and a better communication and comprehension for the user and the listener respectively. Given proven successful techniques to improve the generated speech, the possibility that the speech can be enriched beyond more intelligibility gains motivates the current proposal. Enrichment denotes both modifications to the speech signal itself as well as augmentation.

The general purpose of this project is thus to develop the means to improve spoken language communication throughout the lifespan, primarily in the AAC domain in scenarios such as classrooms, workplaces, household, with a specific focus on enabling speakers and listeners with communication problems to enjoy a greater level of inclusion. In other words, the objective is to pilot the application of speech enrichment in assistive technology.

The proposed work is separated in two steps. The first one consist in finding, understanding and assessing already existing speech generation techniques relevant for the production and improvement of the generated speech on the AAC devices. For example, how to enhance the speech, and reduce the cognitive effort of the listener ? [2] How to generate a speech better revealing the users emotions or personality? [3–5] How to permit the user to be more involved in a communication and be more interesting to listen to? And finally which of all the found techniques are sufficiently advanced to be implemented and used in existing products? Documentation, and contact with the company TOBII Dynavox and other partners from the ENRICH network will enable to list a number of different realistic methods.

In a second step we focus on the most suitable and promising methods and study a way to best adapt and integrate them in the devices with respect to their performance, in terms of data prerequisites (such as availability of speech-,language-, or cognitive models), computational complexity(algorithmic complexity and performance) or required user cooperation.

References

[1] D. J. Higginbotham, H. Shane, S. Russell, and K. Caves, "Access to aac: Present, past, and future," *Augmentative and alternative communication*, vol. 23, no. 3, pp. 243–257, 2007.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[3] C. Nass and K. M. Lee, "Does computer-generated speech manifest personality? an experimental test of similarity-attraction," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2000, pp. 329–336.

[4] M. Schröder, "Emotional speech synthesis: a review." in *INTERSPEECH*, 2001, pp. 561–564.

[5] A. Iida, N. Campbell, S. Iga, F. Higuchi, and M. Yasumura, "A speech synthesis system with emotion for assisting communication," in *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, 2000.