

## Ouisper, Toward An Ultrasound-Based Silent Speech Interface

T. Hueber, G. Chollet, B. Denby, M. Stone

ESPCI-Paristech, ENST-Paristech, U. Pierre et Marie Curie, Paris VI, U. Maryland

The Ouisper project (Oral Ultrasound synthetic SPEech souRce, [1]) proposes to use tongue movements acquired from ultrasound, and lip movements acquired from video to synthesize speech. Such a speech synthesizer, driven only by images, may be qualified as a “silent speech interface” (SSI) and could have several speech applications, such as, use by laryngectomy patients, use in situations where silence must be maintained and augmentation of speech in noisy environments.

The system currently is based on a one-hour audiovisual dataset of ultrasound and video images recorded with the acoustic wave, using HATS [2]. Visual features are extracted from midsagittal ultrasound images using the statistical approach introduced in [3]. This method is a more global approach than classical contour extraction [4], because it uses all the pixels in a 50x50 pixel region of interest and does not extract tongue shape. Instead, it assumes that the ultrasound image can be represented as a linear combination of standard tongue configurations called “EigenTongues,” which are the first Principal Components of the training set. Then, HMM-based stochastic models trained on the Eigentongue decompositions are used to predict phonetic targets from video-only data. Finally, a Viterbi unit selection algorithm is used to find and concatenate the optimal sequence of acoustic units, given this phonetic prediction. The system is already able to perform phonetic transcription from visual speech data with over 50% correct phonetic recognition.

This work is supported by the French Department of Defense (DGA) and the French National Research Agency (ANR).

### References

- [1] Hueber, T., Chollet, G., Denby, D., Stone, M., Zouari, L., 2007. Ouisper: Corpus Based Synthesis Driven by Articulatory Data. ICPhS, Saarbrücken, Germany, to appear.
- [2] Stone, M., and Davis, E., 1995. A Head and Transducer Support System for Making Ultrasound Images of Tongue/Jaw Movement. *Journal of the Acoustical Society of America*, vol. 98 (6), pp. 3107-3112, 1995.
- [3] Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., Stone, M., 2007. Eigentongue Feature Extraction for an Ultrasound-Based Silent Speech Interface. *IEEE ICASSP*, Honolulu.
- [4] Li, M., Kambhamettu, C., and Stone, M. (2005) Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6-7); 545-554.