Consonant Production Control in a Computational Model of Hyper & Hypo Theory (C2H)

Mauro Nicolao, Roger K. Moore
Speech and Hearing Group, Dept. Computer Science, University of Sheffield, UK

Abstract

It is hypothesised that low-energy attractors for the human speech production exist and an interpolation/extrapolation along the key dimension of hypo/hyper-articulation can be obtained by controlling the distance to such an attractor (Moore & Nicolao, 2011). Such hypothesis is evaluated by controlling the consonant production in a speech synthesiser.

A Computational model for H&H theory (C2H) has been developed by connecting a modified HMM-based speech synthesiser and an emulation of human auditory system. The feedback loop evaluates the speech intelligibility of the synthesiser outcome and to control the adaptation. The intelligibility, in terms SII, has been proved to be effectively controlled by using the proposed adaptation.

Motivation

The goal of human speech communication is to transfer information from the talker to the listener minimising the effort and maximising the effectiveness, H&H theory (Lindblom, 1990), paying particular attention to the listener's needs (Moore, 2007).

This experiment is part of a wider project which aims to investigate the possibilities for introducing a feedback path into automatic speech synthesis such that adjustments can be made continuously as a function of perceived effectiveness in the communicative context.

C2H model

The modified synthesiser allows for recursive parameter generation: it handles the auditory-loop information frame-by-frame and adapts the speech production dynamically.

Evaluation

Using C2H, the CPC adaptation was applied to a synthetic voice and the outcome was evaluated. The emulation of auditory feedback was obtained with Speech Intelligibility Index (SII) (ANSI, 1997), a standard index of intelligibility for speech in noise.

As in (Moore & Nicolao, 2011), a set of 200 utterances were synthesised with standard (STD), full-strength hyper- (HYP) and full-strength hypo-articulated (HYO) pronunciation, and compared with the same-intensity and same-quality noise.

Results

Mean values of SII-differences between STD speech and the one modified by maximum-strength direct and inverse CPC adaptation in car noise with $E_{\text{rms}} = -30$ dBFS.

Conclusions

The Consonant Production Control tested in this experiment can be an effective method to modify the intelligibility of synthetic speech in noisy environment. It is worth emphasising that such control emerges from phonetic-contrastalone.

References


Acknowledgement: this research was funded by the European Community’s Seventh Framework Programme under grant n. 213850 - SCALE.

e-mail: m.nicolao@dcs.shef.ac.uk, r.k.moore@dcs.shef.ac.uk