A computer model of perceptual compensation for reverberation based on auditory efferent suppression

Speech and Hearing Group, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield S1 4DP

Amy Beeston and Guy J. Brown
(a.beeston,g.brown)@dcs.shef.ac.uk

Abstract

Watkins (2005) has shown that listeners perceptually compensate for the effects of reverberation by using information from the temporal context preceding a test word. Specifically, Watkins uses a discrimination task in which a test word drawn from a continuum between ‘sir’ and ‘sir’ is embedded in a context phrase, and the reverberation conditions of the test word and context are matched or mismatched. Perceptual compensation (measured by a shift in the category boundary between ‘sir’ and ‘sir’) occurs when the reverberation conditions in the context and test word are consistent.

Reverberation reduces the dynamic range of speech by filling dips in the temporal envelope with energy from reflections. Since the listener’s system has been implicated in the control of dynamic range (Guinan, 2006), we ask whether perceptual compensation for the effects of reverberation could be explained in terms of auditory efferent suppression.

Introduction

- Perceptual constancy allows us to ‘account’ for our surroundings and overcome acoustic distortions likely in natural environments.
- This study presents a computer model that replicates human performance in a listening task concerning compensation for reverberation in the ‘sir/sir’ continuum (Watkins, 2005).
- Gradual repositioning of the temporal envelope of ’sir’ creates the impression of a stop consonant ’t’ in ’sir’. An 11-step continuum of such test words are embedded in a spoken phrase ‘OK next you’ll get’ (TEST) to elicit ‘sir’.

Specs—temporal excitation pattern (STEP)

Semi-closed loop model

A dual resonance non-linear (DRNL) filterbank, originally proposed by Meddis et al. (2007) is used. Efficient suppression is modelled by attenuation in the nonlinear path, as described by Ferry et al. (2008).

Haircell transduction is achieved with a rate-limited scheme (Messing, 2007), and the model of inner hair cell function, giving a representation of auditory nerve firing rate. The firing rates in all frequency channels are summed to give a postulated auditory nerve response which is evaluated with a mean-to-peak ratio metric. This metric assesses the amount of reverberation present and thereby controls the amount of efferent attenuation applied subsequently in the model. A qualitative match to the category boundary for human listeners is found for the sir/sir word discrimination task (Watkins, 2005, experiments 1 and 5). Consistent with human data, the model shows little effect on sir/sir word boundaries under time-reversal of the spoken context phrase, but displays significant disruption of compensation for reverberation under time-reversal of the room reflection pattern.

Specs—temporal excitation pattern (STEP)

Model Framework

The main features of compensation for reverberation are visible in the forward-speech, forward-reverberation case:

- Effect of reverberation: when the test word only is reverberated the category boundary shifts upwards (more ‘sir’ responses). The model is tuned on the forward-speech, forward-reverberation near-near (context-test) and far-far conditions, as the matches Watkins’ data exactly. The remaining conditions also provide a good match to listener data.
- Time reversed speech: compensation for reverberation occurs.
- Time reversed reverberation: compensation for reverberation is abolished.

Results

The features of the model and the parameters are tuned to give good matches with both the speech-processing and perception literature. The model is extended to measure the category boundary shift in the ‘sir/sir’ condition to replicate the human data.