Perceptual compensation for adverse effects of room reverberation on speech recognition: A model based on auditory efferent processing

Amy Beeston and Guy J. Brown

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

{a.beeston, g.brown}@dcs.shef.ac.uk

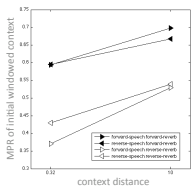
Abstract

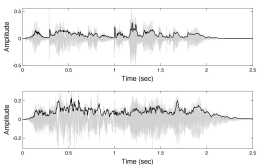
Watkins (2005) found that listeners perceptually compensate for the adverse effects of room reverberation by using information from the temporal context preceding a test word. He embedded a test word from a continuum between 'sir' and 'stir' into a context phrase, and varied reverberation conditions of the context and test independently. Reverberation of the test word alone prompted more 'sir' responses, but similar reverberation of the context permits compensation so that 'stir' responses are restored. An auditory model is described that replicates these effects, finding little change in compensation for reverberation when speech carriers are reversed, but significant disruption of compensation when the reverberation itself is reversed. The simulation suggests that auditory mechanisms controlling dynamic range might contribute to perceptual compensation in the 'sir/stir' task.

Background

- Perceptual constancy allows us to compensate for our surroundings and overcome distortions of naturally reverberant environments.
- In reverberation, dips in a speech signal's temporal envelope are filled with reflected energy and the dynamic range reduces.
- Since the efferent system has been implicated in controlling dynamic range, we ask whether auditory efferent suppression could explain the effects of perceptual compensation for reverberation.

- Real-room reflection patterns were stored as impulse responses (IRs) by Watkins at 'near' (0.32 m) and 'far' (10 m) distances.
- Watkins (2005) imposed the temporal envelope of 'stir' on a spoken 'sir' to create the impression of the stop consonant 't'. An 11-step continuum of such test words was embedded in the context phrase "OK, next you'll get [TEST] to click on".
- Sentences were convolved with IRs to give variously reverberated speech utterances.
- Reflections fill the temporal gap of the 't' in 'stir' making its amplitude envelope similar to that of 'sir' (with reduced dynamic range).
- The category boundary, where 'sir/stir' perception flips, shifts in response to the quality of the preceding context sound.





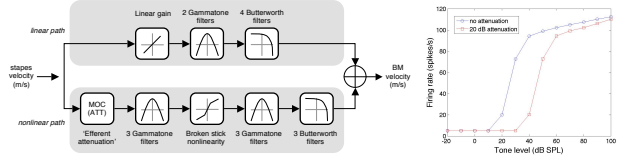
near condition
distance = 0.32 m
MPR = 0.19

far condition
distance = 10 m
MPR = 0.32

- Mean-to-peak ratio (MPR) is tested as a metric to quantify reverberation.
- MPR is inversely proportional to dynamic range, and increases with reverberation.
- MPR is used as a feedback controller to adjust the behaviour of the afferent path of the auditory model depending on the quality of preceding sound.

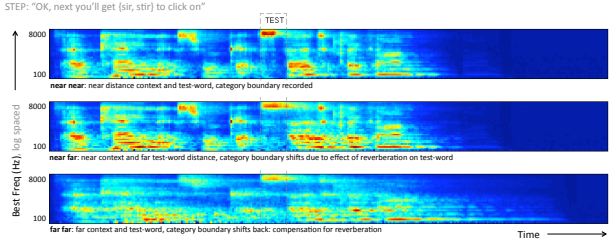
Methodology

- A dual-resonance non-linear filterbank, proposed by Meddis et al. (2001) is configured to represent human listeners (Meddis, 2006).
- As described by Ferry and Meddis (2007), efferent suppression is modelled by varying the attenuation in the non-linear path of the DRNL.
- This helps to recover the dip in the temporal envelope of a reverberated test-word corresponding to the 't' closure in the unreverberated 'stir'.

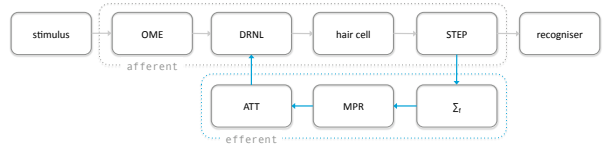


Schematic diagram of a single DRNL filter, redrawn from Ferry and Meddis (2007). Level dependent changes in the bandwidth and centre-frequency are introduced in the lower pathway by the static 'broken stick' nonlinearity. The variable MOC (ATT) refers to the amount of medial olivocochlear attenuation caused by efferent suppression. Increasing attenuation shifts the rate-response curve to higher sound pressure levels.

- Messing (2007)'s model of haircell transduction provides the auditory nerve response.
- Framed, this produces a spectro-temporal excitation pattern (STEP) for recognition.
- STEP templates for 'sir' and 'stir' are stored from the extreme ends of the dry, unreverberated continuum.
- During recognition, these are compared to the STEP resulting from the first 170 ms of the input stimulus test-word (ignoring the vowel), using a minimum mean-square-error distance.



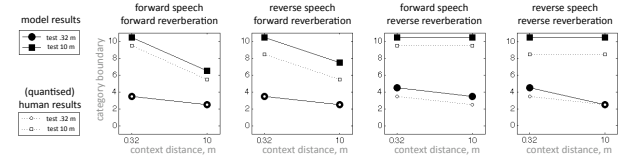
- The STEP is summed across all frequency channels, and the MPR of this summed auditory nerve response is measured over a 1-second time window of the context.
- Continually updating every 1 ms, the derived metric value is recalculated to determine the efferent attenuation (ATT) applied to the DRNL in the next time-step.
- A linear mapping is assumed: both MPR and ATT increase with reverberation.



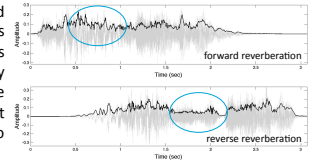
Stimuli are presented to the model's outer-middle ear (OME) at 56 dB SPL, and the afferent pathway subsequently consists of 80 channels with best frequencies from 100 Hz to 8 kHz (log-spaced). The afferent path's output is modulated in a closed-loop feedback system by deriving the next ATT value from the MPR of the pooled STEP.

Results

- The model is tuned on the forward-speech forward-reverberation cases with matching context and test word reverberation (near-near and far-far).
- The remaining conditions provide a qualitative match to listener data in Watkins' study (2005, experiment 5).



- **Effect of reverberation** (in all panels): when the test-word only is strongly reverberated the category boundary shifts upwards (more 'sir' responses).
- **Compensation for reverberation** (panels 1 and 2 only) : when the context is also strongly reverberated the category boundary shifts back down (more 'stir' responses).
- **Reverse speech**: compensation occurs. MPR is little affected by speech direction.
- **Reverse reverberation**: compensation is abolished. The test word is always heard forward, with forward reverberation. Thus when the context reverberation is reversed there is less reverberant energy in the region of the signal preceding the test word. The MPR is reduced, efferent attenuation is reduced, and there is no recovery of the dip in the envelope.



Conclusions

- Model results are consistent with the proposal that auditory processes controlling dynamic range may contribute to the reverberant 'sir/stir' distinction, and that the efferent system may play a role in perceptual compensation for reverberation in this listening task.
- A qualitative fit to human data is obtained when the amount of efferent attenuation is proportional to the mean-to-peak ratio of the across-channel sum of the auditory nerve response in the preceding temporal context window.
- Recent model developments incorporate frequency-dependent attenuation mappings and context 'forgetting' functions that may allow an improved match to human data.

References

Ferry, R.T. & Meddis, R. (2007). A computer model of medial efferent suppression in the mammalian auditory system. *J Acoust Soc Am* 122 (6), 3519-3526.
Guinan, J.J. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear Hear* 27 (6), 589-607.
Meddis, R. (2006). Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. *J Acoust Soc Am* 119 (1), 406-417.
Meddis, R., O'Mard, L.P., & Lopez-Poveda, E.A. (2001). A computational algorithm for computing nonlinear auditory frequency selectivity. *J Acoust Soc Am* 109 (6), 2852-2861.
Messing, D.P. (2007). Predicting confusions and intelligibility of noisy speech. *PhD Thesis*. Massachusetts Institute of Technology.
Messing D.P., Delhorne L., Bruckert E., Braida L.D. & Ghitza O. (2009). A non-linear efferent-inspired model of the auditory system; matching human confusions in stationary noise. *Speech Communication*, 51(8), 668-683.
Watkins, A.J. (2005). Perceptual compensation for effects of reverberation in speech identification. *J Acoust Soc Am* 118 (1), 249-262.

Acknowledgements

This work is supported by an EPSRC grant (EP/G009805/1) entitled 'Perceptual constancy in real-room listening by humans and machines' and is undertaken in collaboration with Anthony Watkins, Simon Makin and Andrew Raimond at Reading University. Thanks to Ray Meddis and Robert Ferry at Essex University for their DRNL program code.