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 ‘stir’ 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 ‘stir’) occurs when the reverberation conditions in the context and test word are consistent.

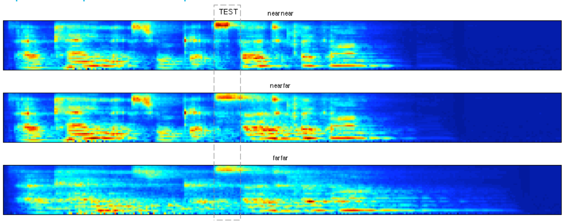
Reverberation reduces the dynamic range of speech by filling dips in the temporal envelope with energy from reflections. Since the efferent system has been implicated in the control of dynamic range (Guinan, 2006), we ask whether perceptual compensation for the effects of reverberation could be explained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

Our present computer model is inspired by the work of Ferry and Meddis (2007) and Messing *et al.* (2009). The acoustic signal is passed through a bank of cochlear filters and a simple 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 pooled 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/stir word discrimination task (Watkins, 2005, experiments 1 and 5). Consistent with human data, the model shows little effect on sir/stir 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 patterns.

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/stir’ continuum (Watkins, 2005).
- Gradual imposition of the temporal envelope of ‘stir’ 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 click on’.
- The *category boundary*, where ‘sir/stir’ perception flips, shifts in response to the quality of preceding sound.
- In real-room reverberation, reflections fill the temporal gap of the ‘t’ in ‘stir’ making its amplitude envelope similar to that of ‘sir’ (with reduced dynamic range).
- The efferent system has been implicated in control of dynamic range (Guinan, 2006). We ask whether it could be involved in compensation for reverberation.

Spectro-temporal excitation pattern (STEP)

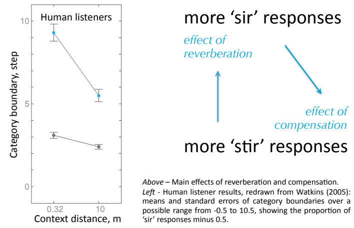


Horizontal axis: time; vertical axis: 80 (log-spaced) frequency channels in the range 100 Hz to 8 kHz.

(i) **near context, near test**: ‘sir/stir’ category boundary recorded

(ii) **near context, far test**: test word only reverberated, category boundary shifts *- reverberation*

(iii) **far context, far test**: context also reverberated, category boundary shifts back *- compensation for reverberation*

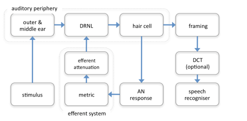


Above – Main effects of reverberation and compensation.
Left - Human listener results, redrawn from Watkins (2005): means and standard errors of category boundaries over a possible range from -0.5 to 10.5, showing the proportion of ‘sir’ responses minus 0.5.

Model

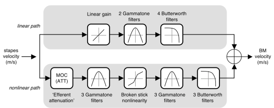
- A dual resonance non-linear (DRNL) filterbank, originally proposed by Meddis *et al.* (2001) is configured to represent human listeners (Meddis, 2006).
- Efferent suppression is modelled by attenuation in the nonlinear path, as described by Ferry and Meddis (2007).
- Haircell transduction is achieved with a rate-limiting scheme (Messing, 2007), and the auditory nerve (AN) response is recorded in a spectro-temporal excitation pattern (STEP).

Model Framework



The model runs 80 channels with best frequencies in the (log-spaced) range 100 Hz to 8 kHz, with stimuli presented at 56 dB SPL. The model is coded in MATLAB, and simulations run on Sheffield’s high performance computing cluster. Features for recognition may be either the STEP (computed at 5 ms intervals over 20 ms windows) or the discrete cosine-transformed (DCT) STEP (15 coefficients, not including the first).

DRNL with efferent attenuation

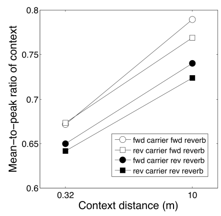


Schematic diagram of a single DRNL filter, redrawn from Ferry and Meddis (2007). A static nonlinearity in the nonlinear path introduces level-dependent changes in the bandwidth and center frequency. The variable ATT refers to the amount of attenuation caused by efferent suppression.

- A semi-closed loop is implemented: context reverberation is judged with a metric, and this determines the amount of efferent attenuation applied thereafter.

Metric: mean-to-peak ratio (MPR)

- The STEP is summed across all frequency channels, and the MPR of this summed AN-response is measured during the 1-second time window of the context sound located directly prior to the test word.
- A linear map is assumed: both MPR and efferent attenuation increase with increasing context distance.
- This mapping helps to recover the dip in the temporal envelope corresponding to the ‘t’ closure in ‘stir’.



Right - MPR of the context, measured over 1 s preceding the test word. MPR increases with context distance. Reversing the speech (carrier) direction has little effect. Reversing the reverberation causes a reduction in MPR relative to the forward reverberation case.

Speech recognition

- STEP templates for ‘sir’ and ‘stir’ are stored from the extreme ends of the dry, unreverberated continuum, and compared (mean-square error distance) with the STEP resulting from the first 170 ms of the test-word in the input stimulus.

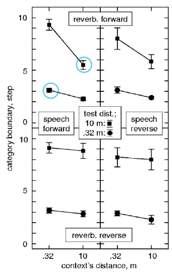
Results

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

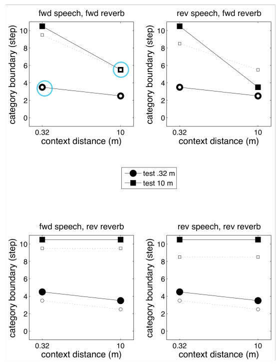
  Effect of reverberation: when the test word only is reverberated the category boundary shifts upwards (more ‘sir’ responses).

  Effect of compensation: when the context and test word are reverberated in the same way the category boundary shifts downwards (more ‘stir’ responses).

- The model is tuned on the forward-speech, forward-reverberation near-near (context-test) and far-far conditions, so these match Watkins’ data exactly. The remaining conditions also provide a good match to listener data.
- **Time reversed speech**: compensation for reverberation occurs.
  The MPR is little affected by the direction of speech, especially at small room distances. The model slightly over-compensates in the far-far condition, but nevertheless provides a qualitative match to human responses.
- **Time reversed reverberation**: compensation for reverberation is abolished.
  There is less reverberant energy in the time window preceding the test-word therefore MPR decreases, less attenuation is applied, and the temporal dip for ‘t’ is not recovered. The match to human results is qualitative, with horizontal lines showing the absence of compensation.



Above – human listener results from Watkins (2005, experiment 5). Blue circles mark the values used to tune the auditory model (right).
Right – human listener results, quantised to possible model values (white); model results (black); tuning points (blue circles). The model predicts qualitatively similar category boundaries to human data. Time-reversal of speech carrier (comparing left with right) shows little effect on sir/stir word boundaries, but time-reversal of reverberation (comparing top with bottom) reveals a significant disruption of compensation for reverberation.



Conclusions

- Model results are consistent with the proposal that the efferent system could play a role in perceptual compensation for reverberation in the ‘sir/stir’ word discrimination task discussed.
- In the semi-closed loop model, a good fit to listener data can be obtained if the amount of efferent attenuation applied to the test word is proportional to the mean-to-peak ratio of the preceding temporal context.
- Auditory processes that control dynamic range may contribute to compensation for reverberation in this listening task.

Ongoing Model Developments

- This is an *across-channel* implementation based on summed auditory nerve response. *Within-channel* mechanisms will be implemented to reflect the frequency-dependency of the efferent system (Guinan & Gifford, 1988).
- Reverberation metrics are under evaluation to implement the model as a closed-loop feedback system with continual update of the efferent attenuation parameter.

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. Jr. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear Hear*, 27(6), 589-607.
Guinan J.J. Jr. & **Gifford** M.L. (1988). Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. III. Tuning curves and thresholds at CF. *Hearing Res,* 37, 29-45.
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. Thanks to Ray Meddis and Robert Ferry of Essex University for the DRNL program code, and additional thanks to Anthony Watkins for supplying experimental stimuli and listener data.

The pilot study for this work (and discussion of an open-loop model) was reported to the Nottingham MRC Institute of Hearing Research workshop *Professional Opportunities for Hearing Scientists* in April 2009. This poster (and discussion of the semi-closed loop model) was prepared for the *British Society of Audiology* Short papers meeting in Southampton, September 2009.