Modelling reverberation compensation effects in time-forward and time-reverse rooms

Amy V. Beeston and Guy J. Brown

Speech and Hearing, Department of Computer Science, University of Sheffield, UK
{a.beeston, g.brown}@dcs.shef.ac.uk

Background

Human listeners
- Human listeners compensate for the effects of reverberation in rooms to overcome distortions.
- In reverberation, dips in a speech signal’s temporal envelope are filled with reflected energy and the dynamic range is reduced.
- Increased reverberation usually degrades recognition performance due to overlap masking (Nábělek, Letenská, & Tucker, 1989).
- If a reverberant test word is presented following a reverberant speech carrier it is more often correctly identified (Watkins, 2005; Brandewie & Zahorik, 2010).
- Compensation for reverberation is incompletely understood.
- The efferent system has been implicated in controlling dynamic range, and shown to be active in noisy environments (Guinan, 2006).
- A monaural compensation effect has been attributed to temporal envelope constancy (Watkins, Raimond, & Makin, 2011; Komada, Bishop, & Klein, 2012).
- Compensation effects are not observed when the reverberation characteristics of the room are time-reversed (Watkins, 2005; Longworth-Reed, Brandewie, & Zahorik, 2008).
- Reﬂected energy prefers the direct sound clearing up prior to onsets.

Machine listeners
- Machine listeners do not typically beneﬁt from a reverberated context.
- Objective measures of reverberation quantify the temporal modulation reduction imposed by a room (e.g., speech transmission index, modulation transfer function).
- These do not depend on the time-direction of reverberation, so cannot explain human listener data.
- A computational auditory model of compensation for reverberation is presented.
- A reverberation metric estimates energy present during tails in the simulated auditory nerve signal.
- This controls an efferent feedback loop which in turn controls the dynamic range of the simulated auditory nerve response.
- Efferent effects have been modelled to improve speech recognition in background noise (Brown, Ferry, & Meddis, 2010).
- We ask whether auditory efferent suppression could explain the effects of perceptual compensation for reverberation.

Auditory Model

- Reﬁnement of model presented in Beeston and Brown (2010).
- Matching conditions in Watkins (2005), stimuli are presented at 48 dB SPL.
- Outer/middle-ear ﬁlter (OME) reproduces resonances of the pinnae and ear canal.
- Simple haircell transduction provides the auditory nerve response (Messing, 2007).
- Spectro-temporal excitation pattern (STEP) templates for ‘air’ and ‘ste’ are stored from the extreme ends of the input un Reverberated 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.
- Efﬁcient suppression is modelled with attenuation (ATT) parameter in non-linear path of the DRNL (Ferry & Meddis, 2007).
- Increasing ATT helps recover dips in the temporal envelope corresponding to stop consonant closures.

Reverberation metrics

Modeling task: sin−air stimuli
- Reverberation-metric (maskshelf) assesses a 1-second context window just prior to test word.
- ATT value for efﬁcient circuit determined by the amount of energy present during reverberant tails.
- Reverse reverberation results in a smaller ATT value; fewer dips are recovered in temporal envelope.

Generalisation: AIC stimuli
- Ideally metric should not conﬂate effect of reverberation with signal content itself. Should also –
  - Be sensitive to source-receiver distance and time-direction of reverberation.
  - Generalise to new stimuli sets and to new rooms.

Stimuli

- Real-room impulse responses (IRs) recorded by Watkins (2005) at six source-receiver distances.
- Sentences convolved with IRs to give various reverberated speech utterances.
- (i) Watkins’ sin−air continuum
  - Temporal envelope of ‘air’ imposed on a spoken ‘sin’ and an 11-step continuum of such test words embedded in the context phrase “OK, next you’ll get [TEST] to click on” (Watkins, 2005).
- (ii) Articulation Index corpus (AIC)
  - Non-sense syllables read by 20 talkers to examine consonant confusions among stops differing in place of articulation: alveolar /t/, velar /k/, and bilabial /p/ (Wight, 2005).
- Reﬂections tilt temporal gaps of stop consonants, reducing dynamic range of the envelopes.
- The likelihood of correct identiﬁcation shifts in response to the quality of the preceding context sound.

Acknowledgements

This work was funded by an EPSRC grant [EP/G009805/1]. Thanks to Tony Watkins, Simon Makin, Andrew Raimond, Kalle Palomäki and Ray Meddis for discussion.

References