* Department of Computer Science, University of Sheffield, UK
§ School of Psychology and Clinical Language Sciences, University of Reading, UK
http://www.personal.reading.ac.uk/~sxs04sjm/constancy

The University Of Sheffield.

University of Reading

Perceptual compensation for reverberation:
human identification of stop consonants in reverberated speech contexts

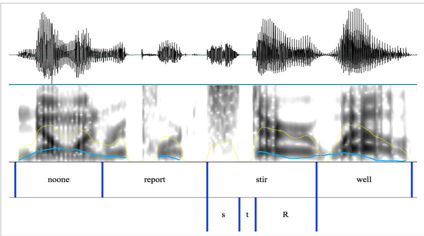
Amy V Beeston* | Guy J Brown* | Anthony J Watkins§ | Simon J Makin§

Background

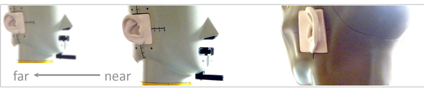
- A speech message is typically understood to have the same phonetic content when heard from a 'near' or 'far' position in a room, despite considerable differences in the temporal envelopes of the speech at these different distances.
- Thus humans show compensation for reverberation (Watkins 2005, Brandewie & Zahorik 2010), just as they exhibit constancy in other ways e.g. for size or brightness.
- Watkins' previous work examined effects of reverberation on the 'sir-stir' distinction by measuring the category boundary position in a continuum of artificially amplitude-modulated stimuli (Watkins 2005, Watkins et al 2010).
- Replicating and extending the 'sir-stir' paradigm, we examine confusions between unvoiced stop consonants differentiated by place of articulation (/p/ front, /k/ middle, /t/ back), using naturally-produced speech with a variety of context words and talkers.

Stimuli

- 80 sentences were selected from the Articulation Index corpus (Wright 2005), each containing the target word 'sir', 'skur', 'spur' or 'stir' (4 targets x 20 talkers).



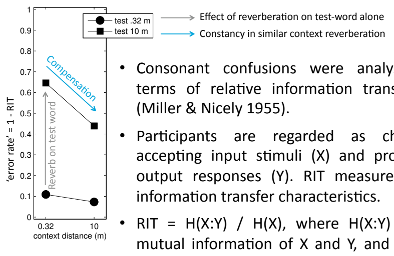
- Target and context were independently convolved with room impulse responses at 'near' (0.32 m) or 'far' (10 m) distances.



- This gave the impression of speech at different positions in a room (an L-shaped office with volume 183.6 m^3).
- Stimuli were presented monaurally (left ear) over headphones in a sound-isolating booth, to listeners who identified the target word.

Constancy effect

- Perceptual compensation for reverberation is observed when increased reverberation on the test word alone caused an increased number of confusions, yet a similarly reverberated context reduced the error rate again.



- Consonant confusions were analysed in terms of relative information transmitted (Miller & Nicely 1955).
- Participants are regarded as channels accepting input stimuli (X) and producing output responses (Y). RIT measures their information transfer characteristics.
- RIT = H(X:Y) / H(X), where H(X:Y) is the mutual information of X and Y, and H(X) is the self-information (entropy) of X.

@nn	sir	skur	spur	stir
sir	58	1	0	1
skur	1	59	0	0
spur	0	0	60	0
stir	0	2	0	58

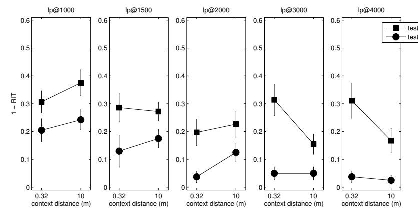
@nf	sir	skur	spur	stir
sir	53	2	1	4
skur	11	47	2	0
spur	11	6	41	2
stir	13	2	0	45

@ff	sir	skur	spur	stir
sir	51	0	0	9
skur	2	52	1	5
spur	1	7	47	5
stir	4	2	0	54

Experiment 1

- Experiment 1 seeks the operating point of this paradigm. To increase the adverse effects of reverberation, experimental stimuli were first low-pass filtered (8th order Butterworth) at 1, 1.5, 2, 3 and 4 kHz cut-off frequencies.
- Matched and mismatched reverberation distance conditions were created: near-near, near-far, far-near and far-far (context-test distance).

- 1600 stimuli (4 targets x 20 talkers x 4 distances x 5 filter conditions) were partitioned among 20 listeners to ensure each utterance was heard once only by each participant.

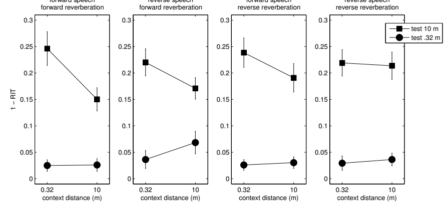


- Errors increase as low-pass filter cut-off frequency decreases: {/k/,/p/,/t/} distinction requires high frequency components.
- Constancy effect is readily apparent only at 3 and 4 kHz low-pass filter cut-off conditions. The 4 kHz condition was selected for further study in Experiment 2.
- A 3-way repeated measures ANOVA (all-within subject factors) found significant main effects of test word distance and filter cutoff frequency, and a significant interaction of test word and context distance.

Experiment 2

- This experiment asks whether time-reversal of the speech or reverberation patterns disrupts the constancy mechanism, as examined by Watkins (2005, experiment 5) for 'sir-stir'.
- The two-word preceding context (cw1, cw2) was reversed in speech direction and/or reverberation direction, while the following context word (cw3) was treated identically to the test word (test).
- Forward reverberation case: preceding context reverberation overlaps test word. Reverse reverberation case: reverberation during test word does not vary with context distance.

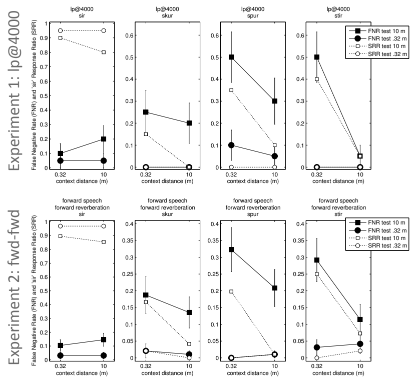
- Stimuli were low-pass filtered at 4 kHz, and four reverberation distance conditions were created as in Experiment 1.
- 1280 stimuli (4 targets x 20 talkers x 4 distances x 4 reversal conditions) were partitioned among 16 listeners. 3 groups of 16 listeners took part (48 participants).



- A 4-way repeated measures ANOVA (all-within subject factors) found significant main effects of test and context distance, and a significant interaction between these two factors.
- Constancy was greatest in the case of forward-speech with forward-reverberation, but reduced when reverberation is reversed. However, the interaction between context distance, test distance and reverberation direction was not significant.

Word-level analysis

- Reverberation caused a particular kind of error to be made in the most typical listening condition (forward speech). Measuring the false negative rate (FNR) per word, and sir response ratio (SRR) as proportion of sir responses per word:
- The majority of errors at near-far were test-words mistaken for 'sir' (SRR≈FNR), as energy from reflections filled in the gap in the temporal envelope that signified /k/, /p/ or /t/.
- At far-far, confusions that persisted in 'skur', 'spur' and 'stir' judgments rarely resulted in 'sir' responses (SRR<<FNR).



Conclusions

- Perceptual compensation for reverberation was apparent in the consonant identification task: increased reverberation on the test word alone caused an increase in confusion rate, but errors were reduced overall when the context was similarly reverberated.
- Listener results from low-pass filtered speech utterances appeared to be consistent with the proposal that constancy may occur band-by-band (Watkins et al 2010).
- The interaction between the constancy effect and the direction of reverberation, as seen by Watkins (2005), was not significantly apparent in corresponding conditions here, possibly due to the higher temporal uncertainty in the present paradigm.
- In typical room conditions, compensation brought about a reduction in mistaken 'sir' responses but confusions persisted between 'skur', 'spur' and 'stir'.

References

- **Brandewie E. & Zahorik P. 2010**. Prior listening in rooms improves speech intelligibility. *J Acoust Soc Am*, 128(1), 291-299.
- **Miller G.A. & Nicely P.E. 1955**. An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am*, 27(2), 338-352.
- **Watkins A.J. 2005**. Perceptual compensation for effects of reverberation in speech identification. *J Acoust Soc Am*, 118(1), 249-262.
- **Watkins A.J., Raimond A. & Makin S.J. 2010**. Room reflections and constancy in speech-like sounds: within-band effects. In: E.A. Lopez-Poveda, A.R. Palmer & R. Meddis (eds.) The neurophysiological bases of auditory perception. New York: Springer, pp. 439-447.
- **Wright J. 2005**. Articulation Index. Linguistic Data Consortium, Philadelphia.

Acknowledgments

Supported by the EPSRC, ref. EP/G01034X/1 and EP/G009805/1. Thanks to Hynek Hermansky and Andrew Raimond.