School of Psychology and Clinical Language Sciences

University of Reading

Room reflections, perceptual grouping and constancy in speech-like sounds

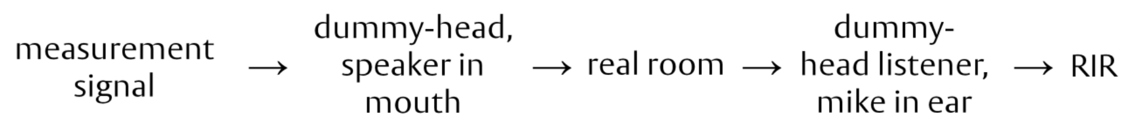
Anthony J Watkins | Simon J Makin | Andrew P Raimond

Background

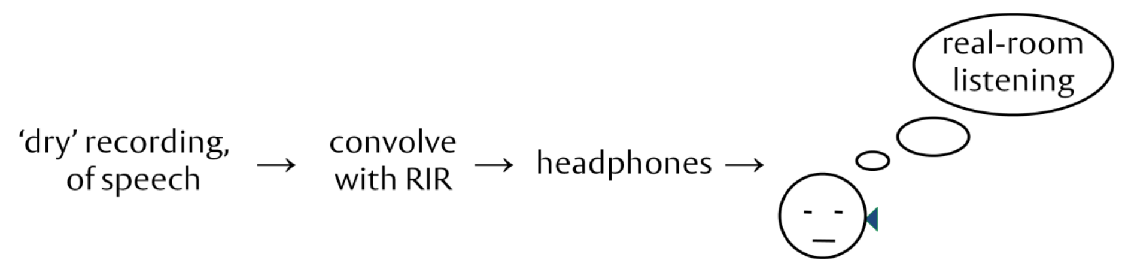
- a speech message played several metres from the listener in a room is usually heard to have much the same phonetic content as it does when played nearby
- however, room reflections make the temporal envelopes of the speech very different at these distances
- this appears to be an instance of 'constancy', due to perception 'taking account' of the level of reflections in neighbouring 'context' sounds (Watkins, 2005a,b)
- here, we measure the effects of this constancy , and ask if it is influenced by different types of perceptual grouping among the context's frequency-bands
- we consider grouping through phonetic factors, as well as grouping through more 'primitive' perceptual factors

Real-room impulse responses, RIRs

- real-room measurements with human-dummy heads, giving room-impulse responses (RIRs) :


measurement signal → dummy-head, speaker in mouth → real room → dummy-head listener, mike in ear → RIR

- RIRs used to effect real-room listening conditions :

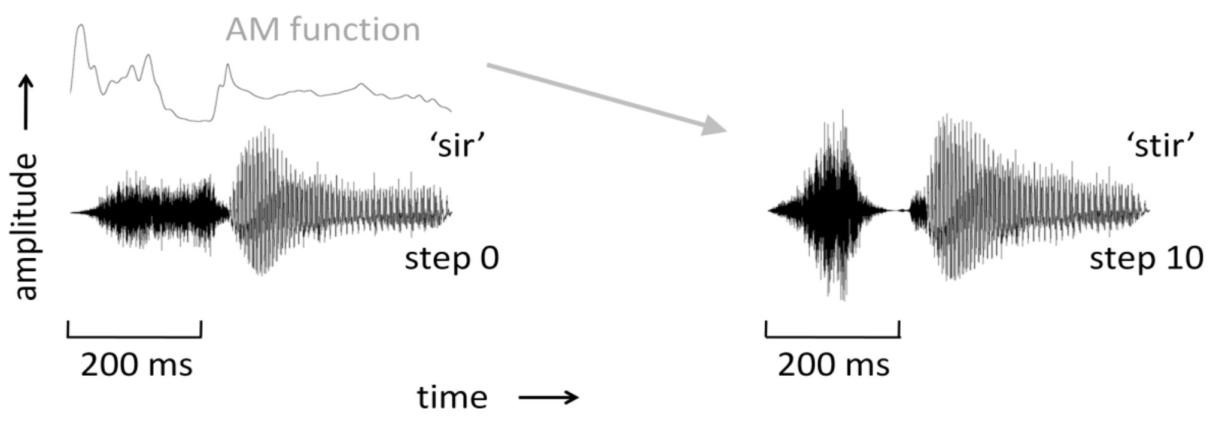

'dry' recording, of speech → convolve with RIR → headphones →

real-room listening

- the level of the room reflections varies with the distance between the heads:
- early (50 ms) to late ratio; 18 dB at 0.32 m → 2 dB at 10 m. (A-weighted energy decay rate; 60 dB per 960 ms at 10 m, room volume = 183.6 m^3)

Test words

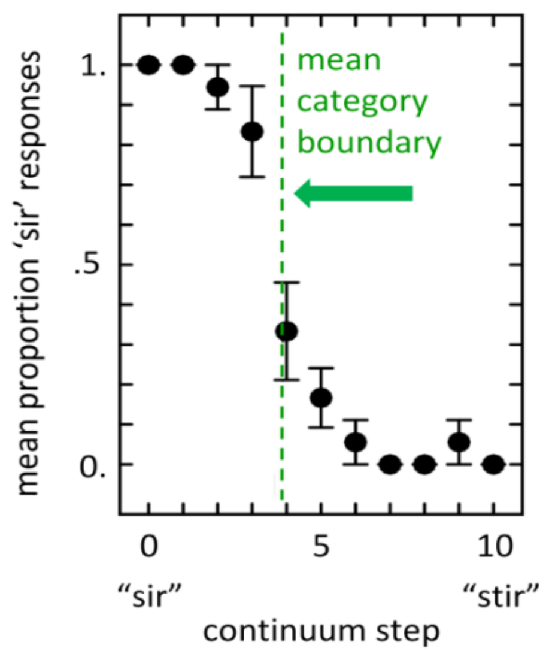
- listeners in 'virtual rooms', hearing RIR-processed sounds
- they identify test words from an 11-step continuum, formed by amplitude modulation (AM) of 'sir', giving 'stir':



- intermediate steps, (1-9) by varying modulation depth

Context and category boundaries

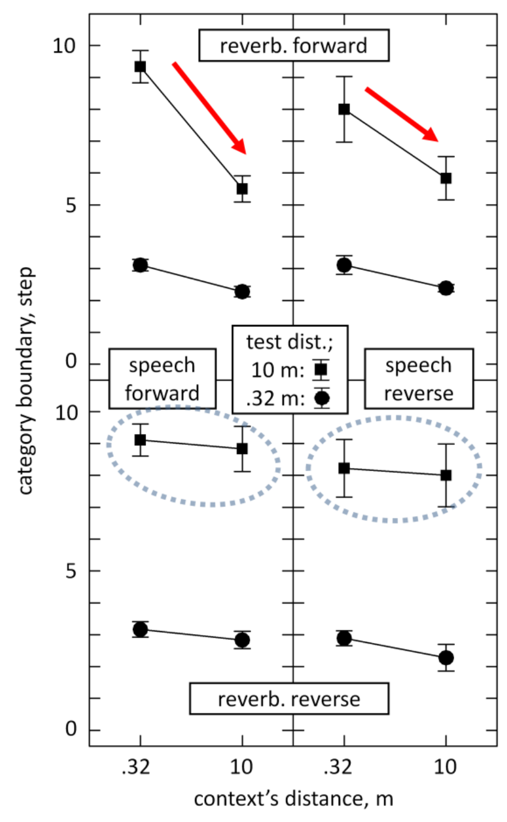
- test-words are played to the listener in the context phrase; 'next you'll get __ to click on'
- listeners respond 'sir ' at lower steps, switching to 'stir' at the higher steps
- this gives a category boundary at the mid-point of the identification function:



Constancy effect

- increase level of reflections (distance) of test sound:
  - more 'sir' responses
  - category boundary increases
- increase distance of context as well → constancy effect:
  - fewer 'sir' responses
  - restores position of category boundary

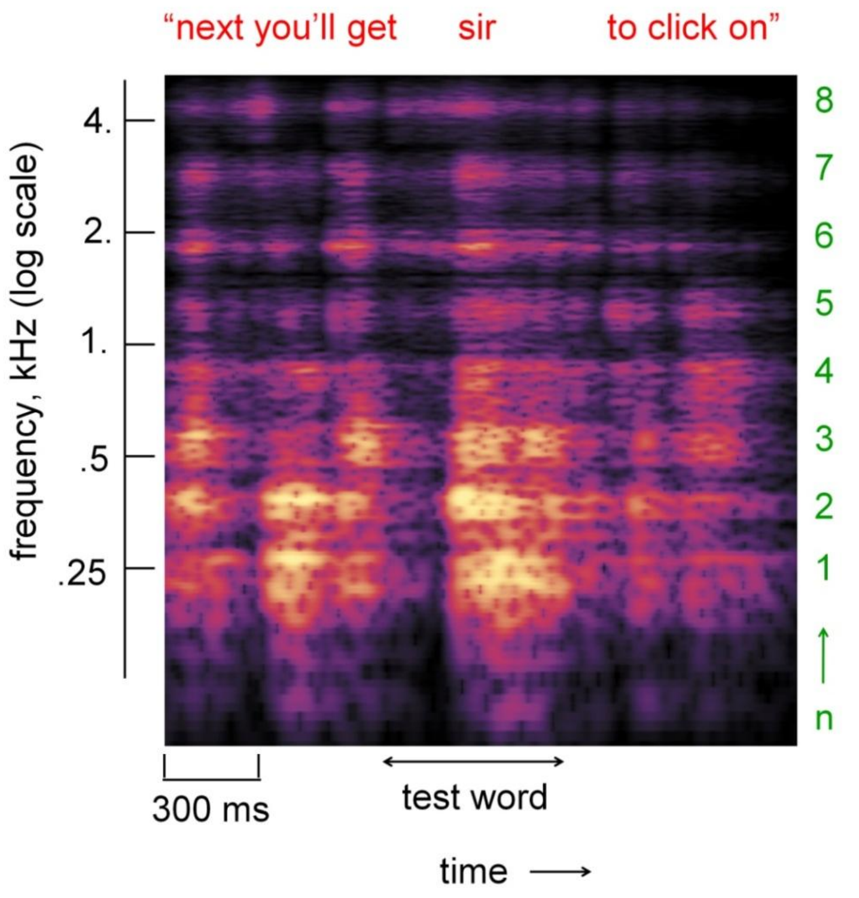
Watkins (2005a) expt. 5



- constancy effect (arrowed) with forwards speech:
- also, a constancy effect (arrowed) when the first and second parts of the context's speech were each played backwards:
- however, when the context's RIR was reversed, giving reversed reverb., the constancy effect was abolished (circled):

Sparse-NV speech and grouping

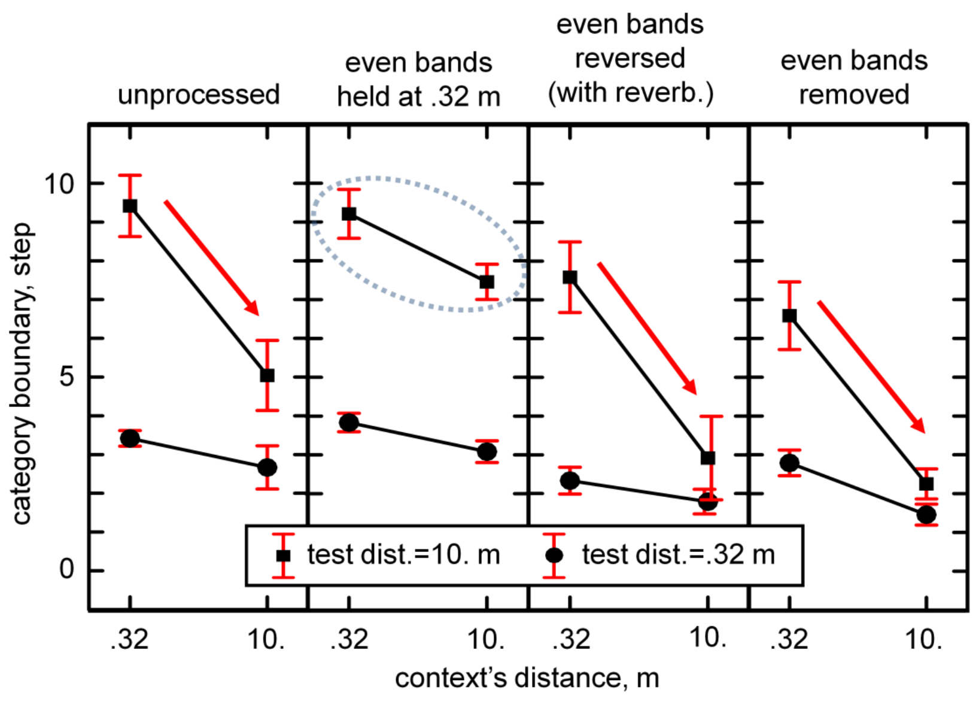
- speech processed with an 8-band noise-excited vocoder
- temporal envelope in each band from gammatone-filtered speech, (η=4, and bandwidths= 'Cambridge ERBs')
- each envelope applied to a (similarly) gammatone-filtered noise
  - n=band number, and n=1,2,...,8
  - band centre-frequencies in kHz = 0.25 x $2^{(7/12)(n-1)}$



- individually, the bands each sound like unintelligible noises
- but when the bands are all played together there is a grouping effect, and the speech-message is heard (Shannon, Zeng, Kamath, Wygonski, and Ekelid,1995)
- here, the effect of reversing only half (4) of the bands is investigated

Experiment 1

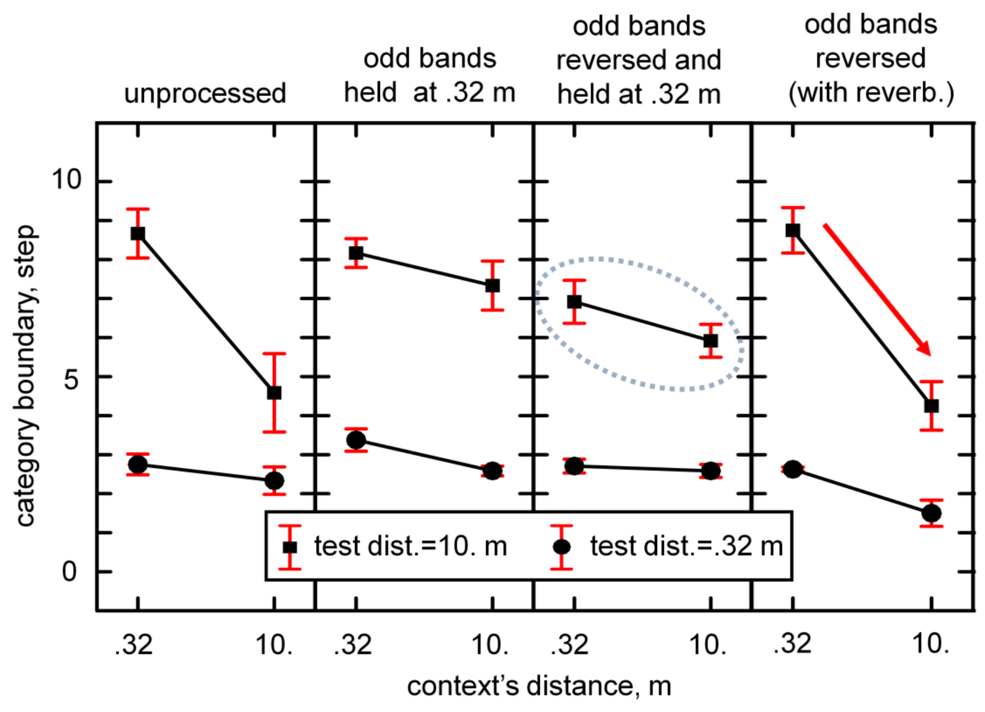
- the context's even-numbered bands were reversed, giving reversed reverb. on speech bands played backwards
- in two other conditions the even-numbered bands were either removed altogether, or their RIR was fixed at 0.32 m



- constancy is reduced (circled) when the even bands are fixed at 0.32 m, presumably because only the other 4 bands are now contributing
- however constancy is substantial (arrowed) in all the other conditions
- hence, the effect of reversing bands in this experiment is similar to the effect of removing them
- this suggests that reversed bands might be grouped separately from the others in perception

Experiment 2

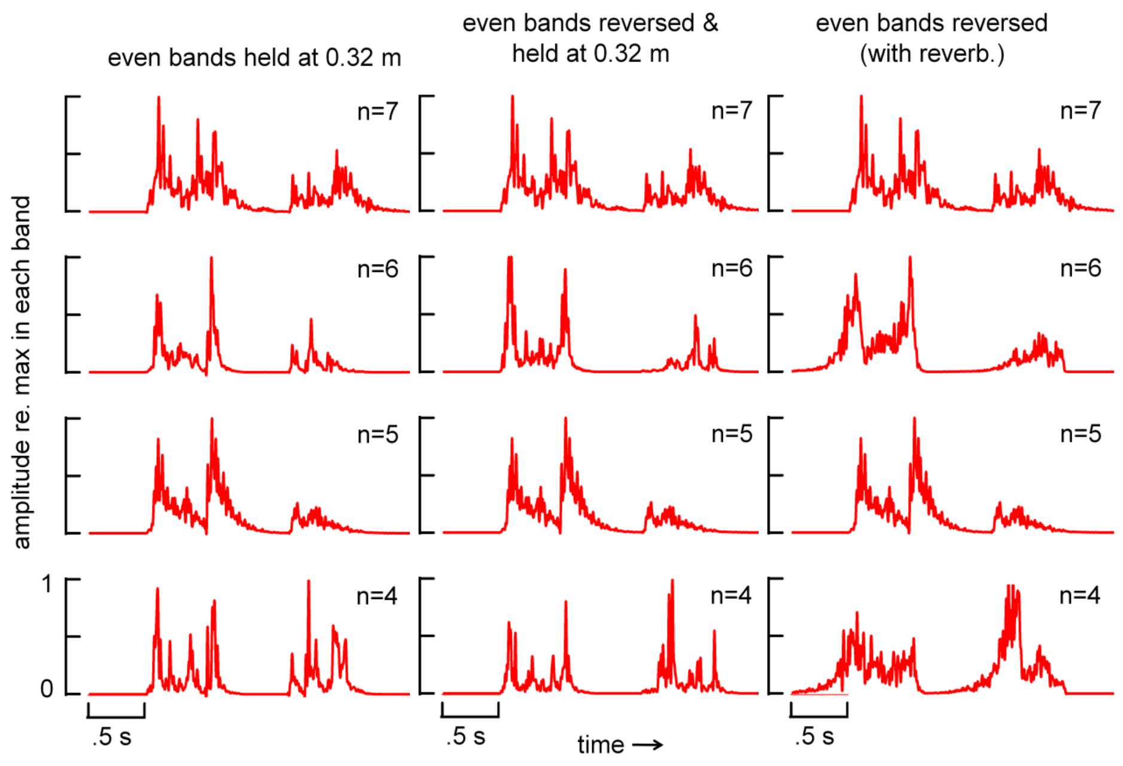
- in forwards conditions, concurrent sounds are all associated with the same phoneme, but in reversed conditions this is not the case
- does this phonetic factor give the different groupings seen in experiment 1?
- here, the context's odd-numbered bands are manipulated, and in different conditions they are reversed, held at 0.32 m, or both
- both of the reversed conditions should give substantial constancy if phonetic factors are effecting a grouping



- constancy is substantial when the speech-band and its reverb. are played backwards (arrowed)
- by comparison, constancy is much less substantial when the speech band is played backwards and the reverb is at 0.32 m (circled)
- so the reversing effect observed in experiment 1 replicates when the odd-numbered bands are reversed, but the groupings responsible do not seem to involve a phonetic mechanism

Discussion

- the temporal envelopes in 4 of the context's bands are shown below
- 'primitive' grouping cues can be seen by comparing bands that have reversed reverb. with bands that have forwards reverb., particularly at onsets



Conclusions

- the grouping of bands in NV speech appears to arise from mechanisms more primitive than those responsible for phonetic perception
- nevertheless, the speech-like phonetic quality of these sounds seems to arise from this primitive grouping
- mechanisms of perceptual constancy seem to precede this grouping

References

1. Shannon, R.V, Zeng, F. Kamath, V. Wygonski, J. and Ekelid, M. (1995) Speech recognition with primarily temporal cues. *Science* **270** 303–304
2. Watkins, A.J. (2005a) Perceptual compensation for effects of reverberation in speech identification. *J. Acoust. Soc. Am.* **118** 249-262
3. Watkins, A.J. (2005b) Perceptual compensation for effects of echo and of reverberation in speech identification. *Acta acustica united with Acustica* **91** 892-901

Acknowledgements

- Thanks to Amy Beeston, Guy Brown, Peter Derleth , Kalle Palomäki and Hynek Hermansky for discussions
- Supported by an EPSRC grant to the first author

Further information

www.reading.ac.uk/~syswatkn