

Perceptual constancy, reverberation, and grouping: Within-band effects

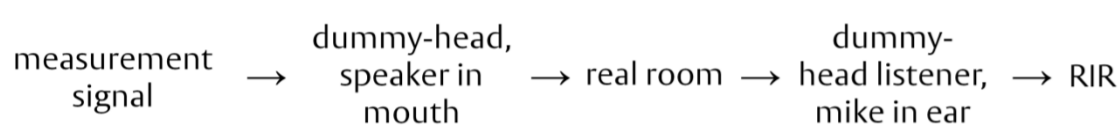
Anthony J Watkins | Simon J Makin | Andrew P Raimond

Background

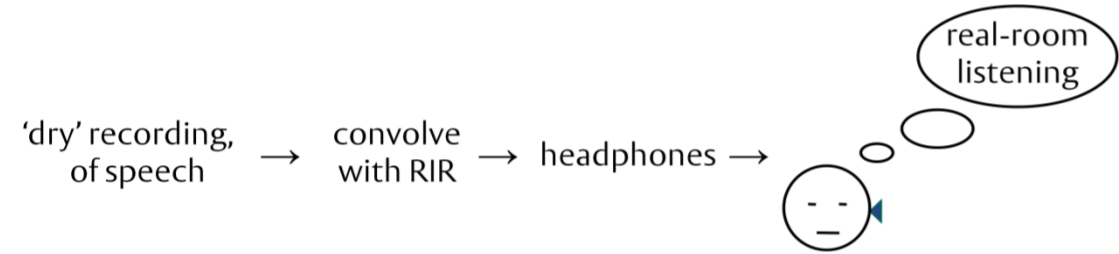
- a speech message played several metres from the listener in a room is usually heard to have much the same phonetic content as it does when played nearby
- however, room reflections make the temporal envelopes of the speech very different at these distances
- this appears to be an instance of 'constancy', due to perception 'taking account' of the level of reflections in neighbouring 'context' sounds (Watkins, 2005a,b)
- recently, vision researchers have asked where constancy stands in relation to perceptual grouping (Palmer, Brooks and Nelson 2003)
- the present experiments ask the same question of hearing

Real-room impulse responses, RIRs

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



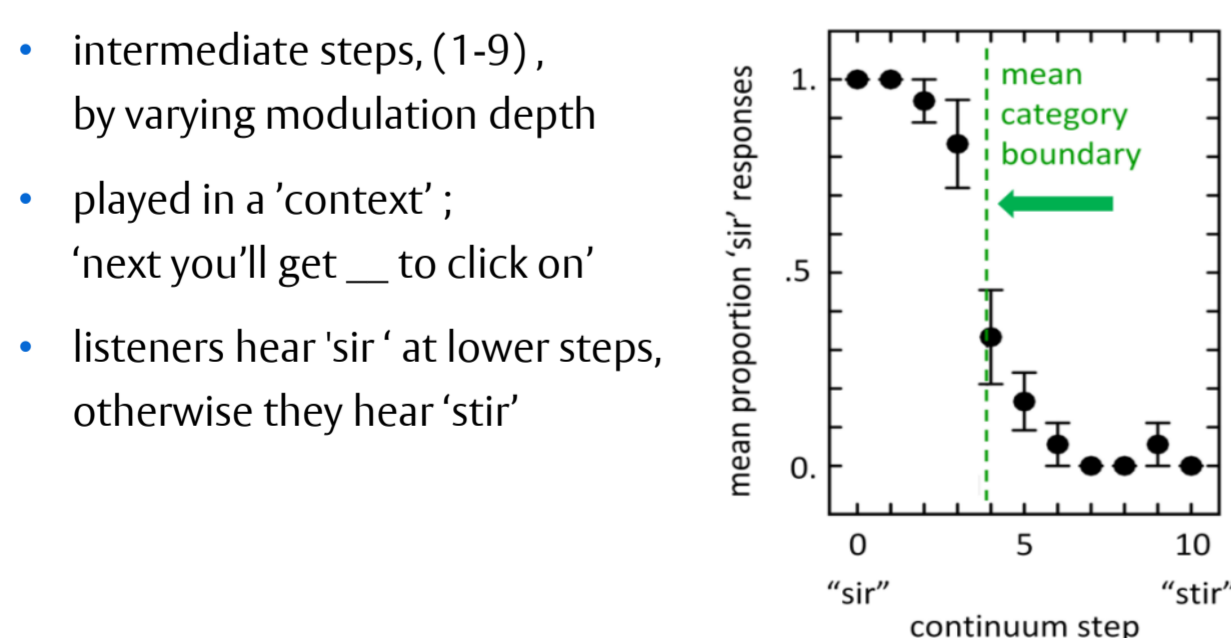
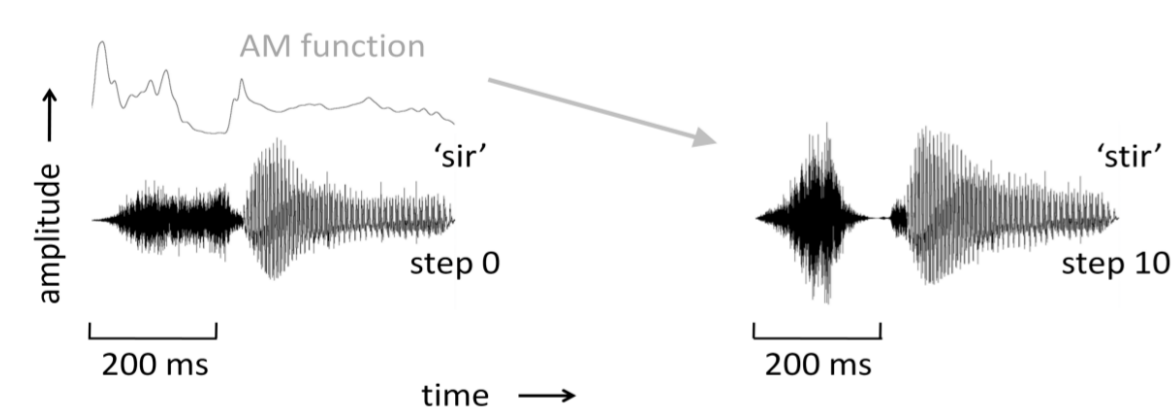
- RIRs used to effect real-room listening conditions:



- the level of the room reflections varies with the distance between the heads:
- the distance between the heads is used here to vary the level of room reflections:
 - 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, and room volume = 183.6 m³)

Test words and category boundaries

- 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':



Sparse-NV speech

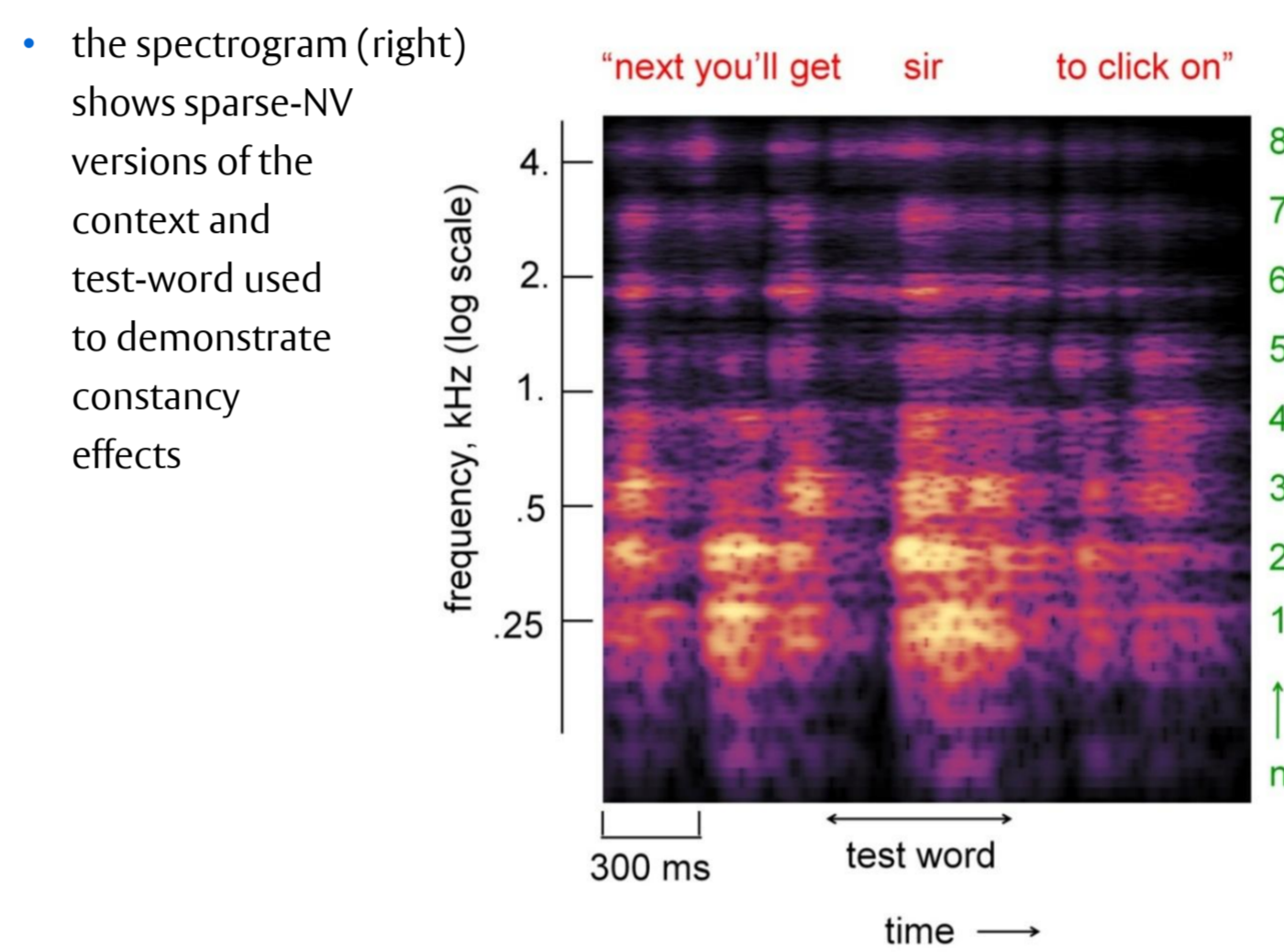
- speech processed with an 8-band noise-excited vocoder
- temporal envelope in each band from gammatone-filtered speech, ($\eta=4$, bandwidths= 'Cambridge ERBs')
- each envelope applied to a (similarly) gammatone-filtered noise
 - n =band number, and $n=1,2,\dots,8$
 - band centre-frequencies in kHz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

Grouping & sparse-NV speech

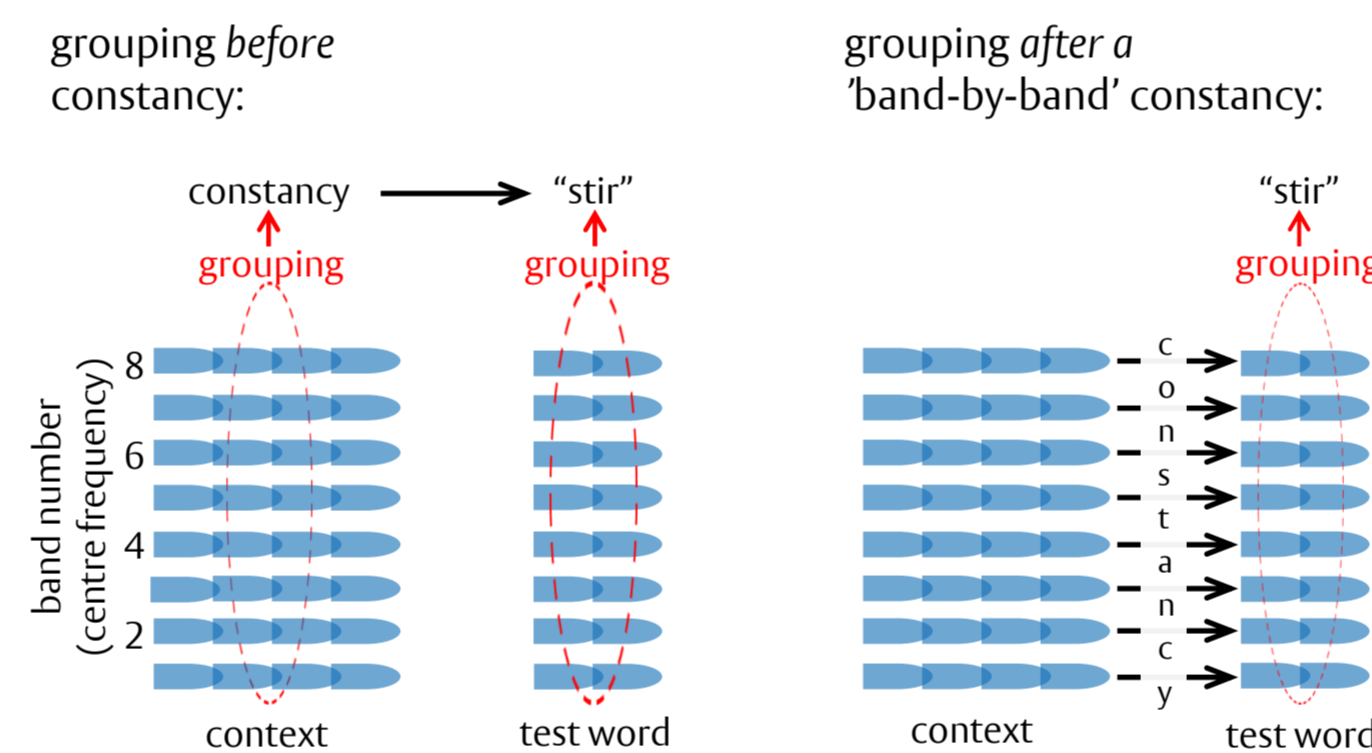
- individually, the vocoder's bands each sound like unintelligible noises
- but when the bands are all played together there is a grouping effect, and the speech-message is heard (Shannon, Zeng, Kamath, Wyganski, and Ekelid, 1995)

Constancy & sparse-NV speech

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

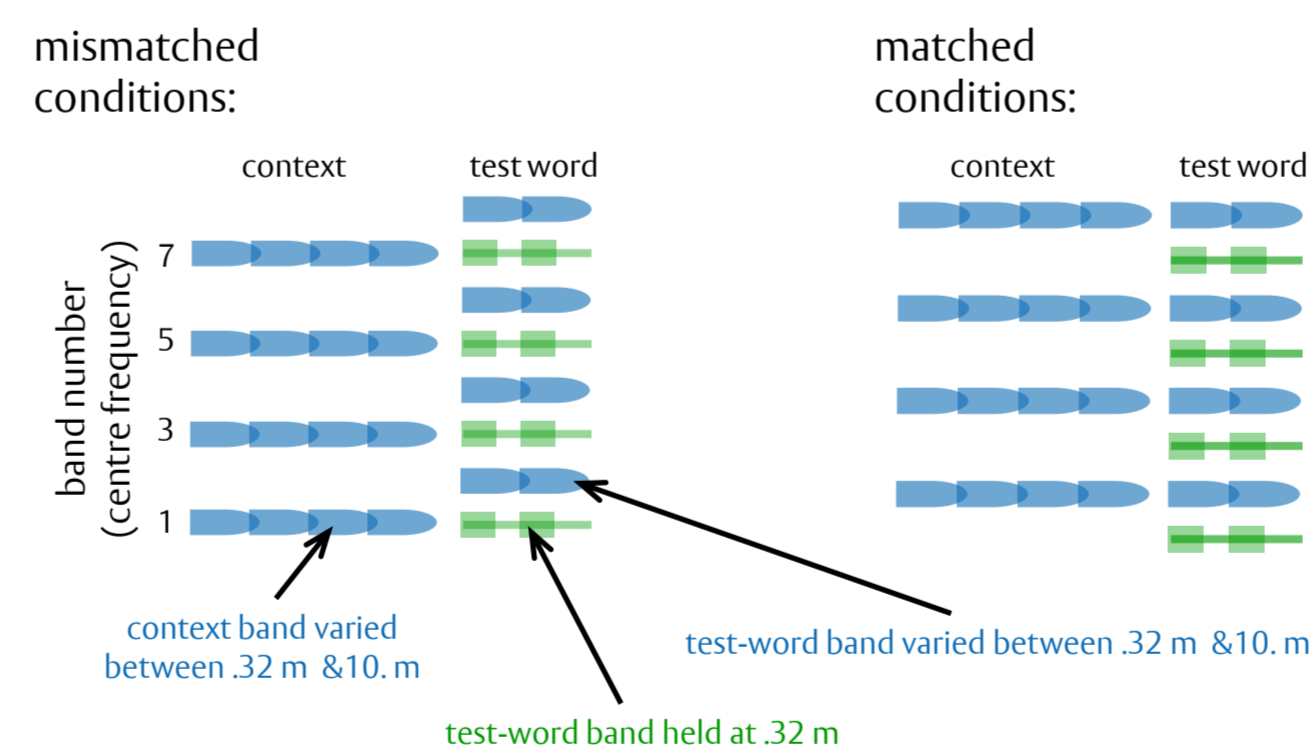


When does grouping happen?

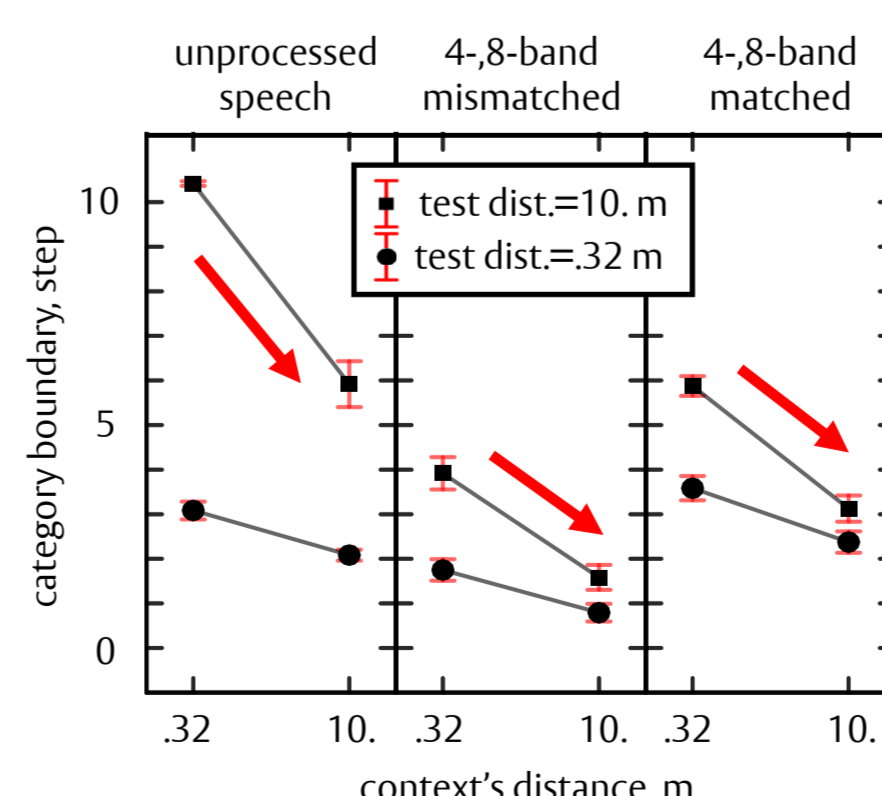


Experiment 1

- 4-band contexts and 8-band test words
- distance varied in the context, but only in 4 of the test-word's bands:

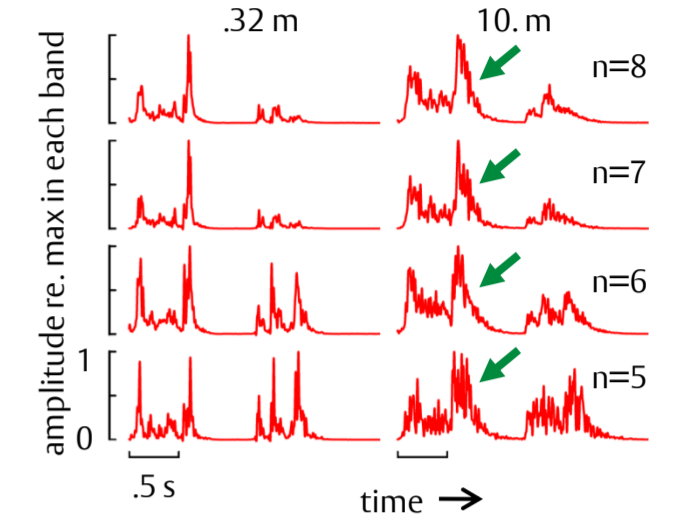


- is the constancy effect reduced in mismatched conditions?
- no, there are constancy effects (arrowed) in all conditions:



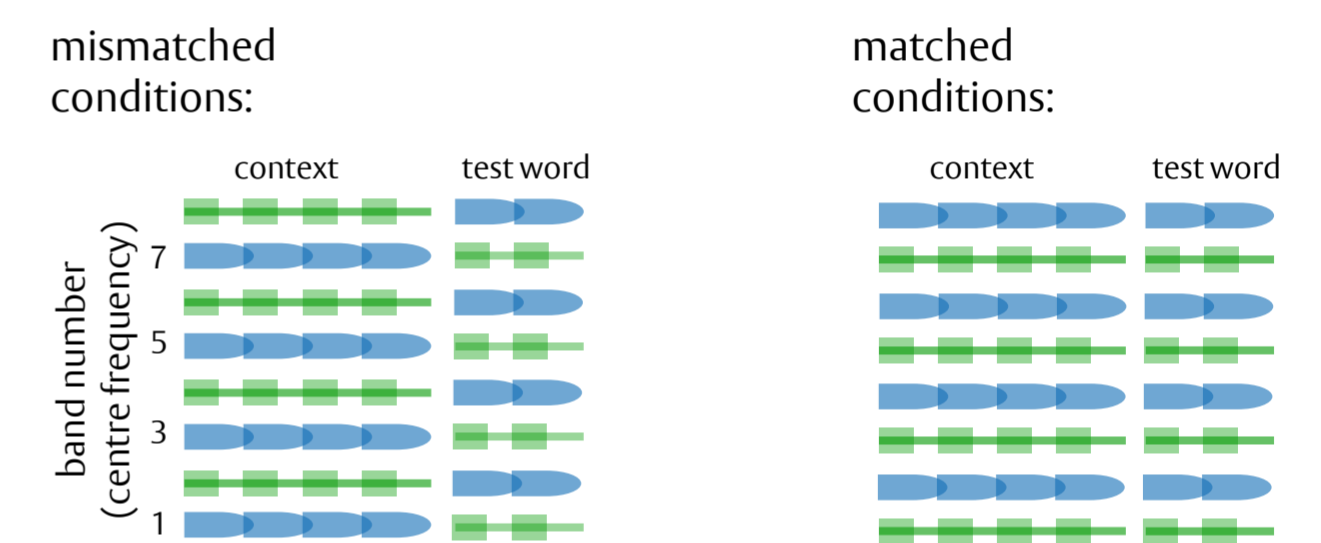
Spread of excitation and 'tails'

- the context's temporal envelopes in auditory (gammatone) filters are shown on the right:
 - in the more distant, 10-m sounds, there are prominent 'tails', and examples are arrowed
 - in these mismatched conditions, the temporal envelopes in both odd and even-numbered bands of the contexts show these tails, even though the even-numbered bands are nominally missing
 - this 'cross-band leakage' of tails seems to arise through a spread of excitation, which experiment 2 tries to eliminate

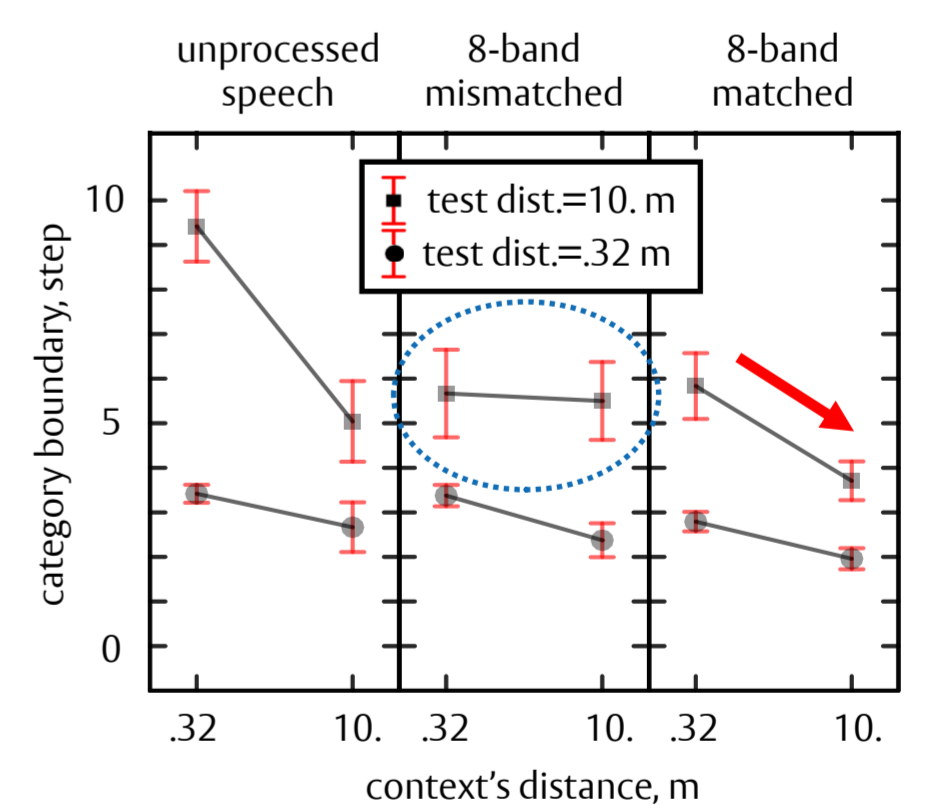
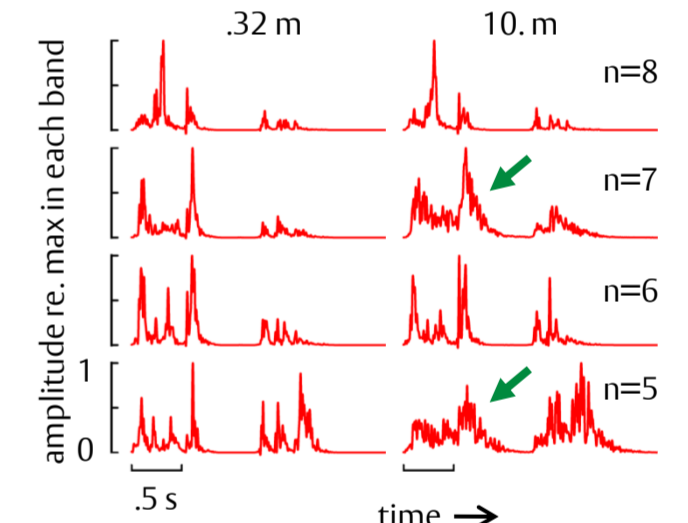


Experiment 2

- 8-band contexts and 8-band test words
- distance varied in only 4 of the test word's and the context's bands:



- the context's temporal envelopes in expt. 2's mismatched conditions are shown on the right:
 - now, prominent tails are only seen in the context's odd-numbered bands
 - as a result, constancy is eliminated in mismatched conditions (circled), but it remains in matched conditions (arrowed):



Conclusion

- constancy is effected 'band-by-band' through a mechanism that operates within frequency bands but not between them
- this operation precedes perceptual grouping across the bands

References

- Palmer, S.E. Brooks, J.L. and Nelson, R. (2003) When does grouping happen? *Acta Psychologica*, 114 311-330
- Shannon, R.V., Zeng, F., Kamath, V., Wyganski, J. and Ekelid, M. (1995) Speech recognition with primarily temporal cues. *Science* 270 303-304
- Watkins, A.J. (2005a) Perceptual compensation for effects of reverberation in speech identification. *J. Acoust. Soc. Am.* 118 249-262
- 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