School of Psychology and Clinical Language Sciences

University of Reading

The dominance of ILD cues when tracking talkers in real-room reverberation

Simon J Makin | Anthony J Watkins | Andrew P Raimond ({s.j.makin, syswatkn, a.raimond}@rdg.ac.uk)

Introduction

- Listeners can selectively attend to ('track') a single talker in the presence of other, simultaneous talkers, even in reverberant conditions
- Diverse signal characteristics can help listeners track a speech message over time (e.g. , a filtering difference; Spieth and Webster, 1955)
- Two main sources of such 'tracking cues' in real rooms:
  1. Cues from differences in spatial position such as the interaural time and level relationships (ITD & ILD, Broadbent, 1954; Darwin and Hukln, 2000)
  2. Cues from differences in talker characteristics such as pitch and vocal-tract size (Darwin and Hukin, 2000)
- Interaural cues are corrupted by reverberation in rooms (Kidd et al., 2005), whereas talker-difference cues are very resistant to reverberation (Darwin and Hukin, 2000)
- So, in real-room talker-tracking, do listeners simply ignore the corrupted cues from spatial position and rely on cues from talker differences?

Experimental paradigm

- Listening task where cues from talker- and spatial-differences are in conflict, and where listeners' responses indicate which cue they're tracking
- Based on Darwin and Hukin's (2000) paradigm, where listeners hear two simultaneous messages played in a (simulated) room:

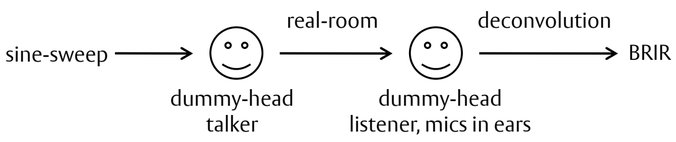
Target sentence: "**On this trial you'll get the word <> to select"**

Distractor sentence: "**You'll also hear the sound <> played here**"

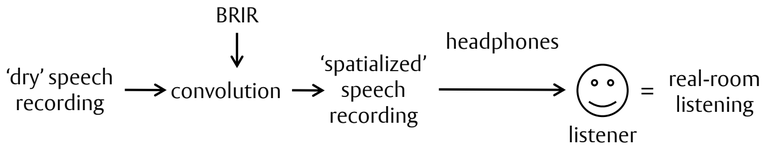
- Recorded with 1 female and 2 male talkers, along with two test words: "**bead**" and "**globe**", which were spliced into the <> position and time-aligned
- Messages and test words were individually 'spatialised ' to vary cues from spatial position-differences
- Listeners were asked to attend to the target message and report which test word they perceived as belonging in it

Real-room spatialisation

- Binaural Room Impulse Responses (BRIRs), recorded in a room using the swept-sine method (Farina, 2000):



- BRIRs were used to spatialise 'dry' speech recordings:

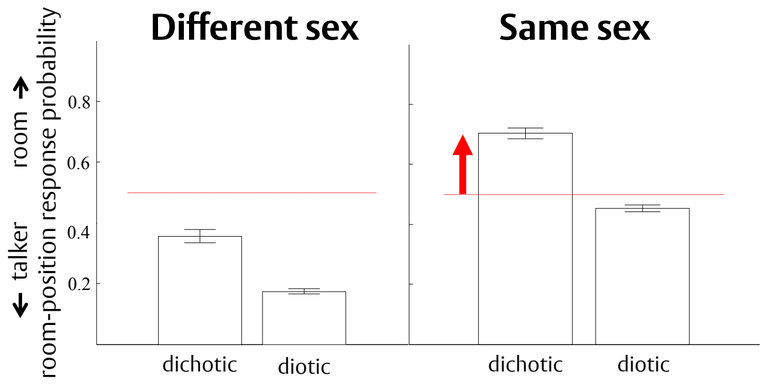


- Corrections were applied for the frequency responses of the headphones and dummy-head talker

Design

- Diverse room-position pairs were used, varying the distance between listener and talkers (0.65, 1.25, 2.5 and 5m), the bearing separation between talkers (+/- 25^0 and +/- 5^0), and both (e.g., a 0.65 m distance at $+5^0$ with a 5 m distance at -5^0)
- Listening was either dichotic, or diotic with the L or R channel presented to both ears and matched to the dichotic level
- The dependent variable was the probability of a room-position response from a listener, averaged across all room-position pairs

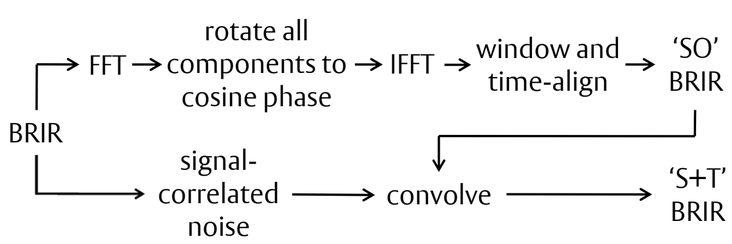
Results: Talker differences



- Talker-difference tends to dominate over room-position, particularly for the different-sex pair
- However, room position is not always ignored, and can be dominant for same-sex pairs ***in dichotic conditions***

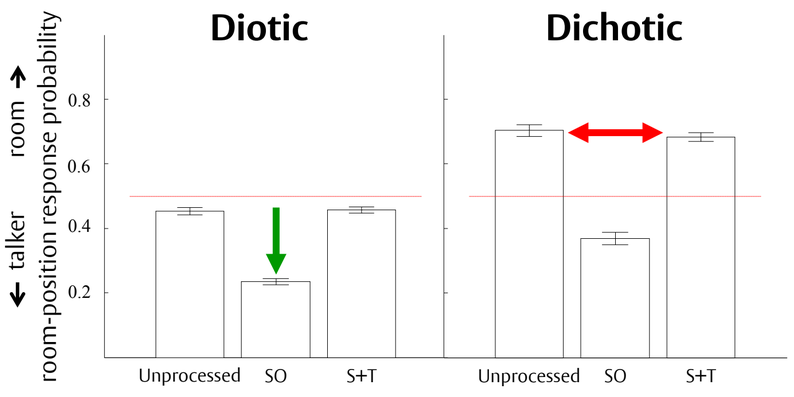
BRIR processing

- To investigate which cues from position are responsible for this dichotic effect, the BRIRs were processed to limit the cues available to listeners as follows:



- Spectral-Only ('SO') BRIRS:
  – all ITDs and temporal-envelope 'tails' are removed
  – leaves only spectral-envelope and level (e.g. ILD) info.
- Spectral-plus-Temporal-envelope ('S+T') BRIRS:
  – as 'SO' but with 'tails' restored

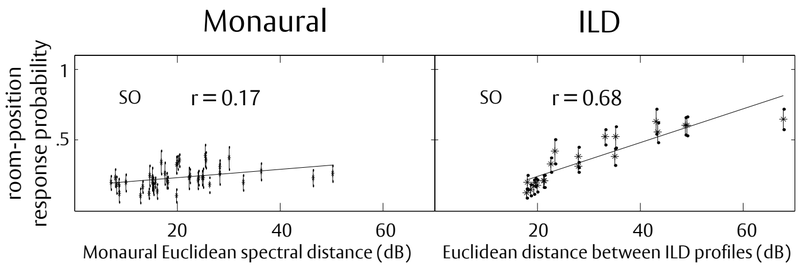
Results: Processed BRIRs



- Talker dominates in diotic conditions, particularly when there are no 'tails'
  – suggests 'tails' disrupt pitch cues (Culling et al., 1994)
- Room position dominates in dichotic conditions, **even in the absence of ITD cues**

Spectral distance between BRIRs

- In dichotic conditions, when listeners track by room-position, do they 'select' the ear with the bigger spectral-distance between the two messages?
- An 'auditory' (gammatone) filter-bank analysis allows calculation of *monaural* spectral distances (d) between the BRIRs in an ear (Euclidean distance, d = the rms of the pairwise dB differences in frequency channels)



- In SO, *diotic* conditions, there is only a weak correlation between monaural spectral distance and room-position response probability
- So ear selection on the basis of these monaural distances seems unlikely to be the origin of the dichotic effect

Spectral distance between ILD profiles

- In dichotic conditions, do room-position responses increase with the ILD difference between the two messages?
- Subtracting a BRIR's L-channel filter-bank spectrum from its R-channel spectrum gives an ILD 'profile' for that position
- The ILD-difference of the two messages is therefore the rms of the differences between their ILD profiles
- These ILD-differences are strongly correlated with the probability of a room-position response ($r^2 = 0.46$)
- So the dichotic effect seems to arise through *inter*-aural processing ***of ILD***

Conclusions

- When tracking a talker, cues from talker differences are not always dominant. Cues from position differences can sometimes be more influential - even in reverberation
- This is mostly seen when talker differences are subtle, listening is dichotic, and pitch cues are degraded
- The cues from position differences are not the messages' ITDs. They seem to be the ILDs - which still differ among positions in a typical room
- This dichotic effect doesn't seem to arise through listeners 'selecting' an ear – it seems to be due to inter-aural processing

References

1. Broadbent, D. E. (1954) The role of auditory localization in attention and memory span. *J. Exp. Psych.* **47** 191-196
2. Culling, J. F., Summerfield, Q. and Marshall, D. H. (1994) Effects of simulated reverberation on the use of binaural cues and fundamental frequency differences for separating concurrent vowels. *Speech Commun.* **14** 1508-1516
3. Darwin, C. J. and Hukin, R. W. (2000) Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *J. Acoust Soc. Am.* **107** 970-977
4. Farina, A. (2000) Simultaneous measurement of impulse response and distortion with a swept-sine technique. 108th AES Convention, Paris, 18th-22nd Feb, 2000
5. Kidd, Jr., G., Mason, C. R., Brughera, A. and Hartmann, W. M. (2005) The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acust. Acta Acust.* **91** 526-526
6. Spieth, W. and Webster, J. C. (1955) Listening to differentially filtered competing voice messages. *J. Acoust Soc. Am.* **27** 866-871