Introduction

Listeners can selectively attend to a desired talker in the presence of interfering talkers, and a spatial position difference is known to aid this ‘tracking’. For instance, much previous research has demonstrated the effectiveness of cues arising from a difference in bearing between the signals (e.g. Bronkhorst and Plomp, 1988).

But there is reason to question the utility of such cues in real-room reverberation, as it degrades interaural localisation cues (e.g. Kidd et al., 2005). Other aspects of position in a room may play a part however. For instance, the decrease in Direct-to-Reverberant energy Ratio (DRR) with increasing distance between listener and source is a cue to perceived distance (Zahorik, 2002). More generally, reflections distort sounds (spectral- and temporal-envelopes), and this varies with position, potentially giving rise to position-specific ‘grouping’ cues.

Listeners can also track talkers using differences in their voice characteristics (Darwin and Hulin, 2000), so this study asks how effectively cues arising from position (where both bearing and distance are varied) compete with cues arising from talker differences, while listening in real-room reverberation.

Methods

Experimental paradigm

• A selective attention paradigm devised by Darwin and Hulin (2000) was used, where subjects hear two simultaneous sentences played in a (simulated) room.

• Target sentence: ‘On this trial you’ll get the word < > to select’

Distractor sentences: ‘You’ll also hear the sound < > played here’ recorded by two male talkers, along with two test words: (’bead’ and ‘globe’), which were spliced into the < > position and time-aligned to be simultaneous.

• Listeners were asked to attend to the target sentence and indicate which test word they perceived as occurring in it. The two sentences and test words where individually spatialised such that talker and spatial attributes were in conflict. Listener response thus indicates which cue is more compelling.

• Listener’s position was fixed. Speakers were placed at two locations in each condition, as shown:

bearing + delta distance

bearing only

Fixed at 5 m

delta distance

distance (meters)

Listener

Listener

Distance was varied (0.65, 1.25, 2.5 and 5 m), while equating overall room levels (taking both channels together, so ILDs preserved) to eliminate level cues to distance, and two bearing differences (+25, 25 and 45°) were used.

• Mean probabilities of room-position response from 8 listeners.

• Stimulus were presented both dichotically and stastically to assess the contribution of binaural hearing. Dichotic stimuli were the LR channel presented to both ears and were level corrected to match dichotic stimuli.

Real-room reverberation

• Stimuli were spatialised using Binaural Room Impulse Responses (BRIRs), recorded using the swept-sine method (Farina, 2000).

• BRIRs were processed to limit the cues available, as follows:

- To retain only ILD and spectral-envelope cues:

- To add temporal-envelope cues:

- Auditory spectra (dB, re:1) obtained by processing BRIRs with auditory model (32 channel filterbank; equally spaced in DRR space). Solid lines are spectra for the +25° bearing, dashed lines for -25°, and dotted lines for 0°. Top row shows spectra for the left channel, available to the right ear, and bottom row is the right-symmetric filterbank.

- Note the distinctly different spectra for each position, and the distinct patterns of ILD variation in dichotic conditions, which could act as ‘timbre signatures’

• Cues to location in real-rooms

Classical localisation cues

• Different angles of incidence give rise to differences in the interaural time delays (ITD) and frequency-dependent interaural level differences (ILD) of the signals, leading to binaural interaction effects.

• Also, two sounds originating from different bearings will be differentially affected by head movements, and one will have a higher signal-to-noise ratio (SNR) in one ear, and vice-versa. Selectively attending to one or the other can thus effectively increase the SNR (so-called ‘better ear listening’). The ear with the better SNR could thus act as a cue to position without recourse to truly interaural processing.

Temporal-envelope distortion

• The amount of reverberant energy in a sound, relative to direct sound energy, increases with distance from source to listener, so that as a sound moves farther away in a room, its temporal-envelope is increasingly distorted by the tails which multiple reflections create at offsets:

\[ r(t) = \sum_{n=1}^{N} h_n(t) \cos(\omega t + \phi_n) \]

- Such distortion is thus position-specific and could aid selective attention by functioning as a ‘timber difference’ grouping cue.

Spectral-envelope distortion

• RIRs from different room-positions have different frequency responses. Convolving with the RIR will superimpose this frequency response on the speech spectrum, thus creating position-dependent ‘colouration’ of the long-term average spectrum:

\[ \text{Spectral-envelope distortion} = \text{RIR} \ast \text{Spectrum} \]

• Also, in order to assess the role of ‘better ear listening’ as opposed to truly binaural processing, a ‘better ear’ analysis was conducted by using an auditory model to calculate Euclidean distances between the spectra of the two RIRs at each ear, with the ear receiving the most different signals in the dichotic condition then used in the better-ear analysis of the corresponding distratic data.

Results

Listeners’ responses with real-room BRIRs

- In dichotic conditions room-position is dominant. The influence of position is diminished in dichotic conditions, but room-position can still be the dominant cue if there is a distance separation between talkers.

- There are prominent effects of distance separation in the dichotic data, particularly when the bearing separation is small. It’s possible this is due to temporal-envelope cues as the length of ‘tails’ in the temporal-envelope increases with distance.

Processed BRIRs: ‘SO’

- In dichotic conditions room-position is still dominant. This can only be due to ILD and spectral-envelope effects. There is no ITD information, and no ‘better ear’ effect so the large binaural advantage seems mainly due to ILD. No longer any extra effect conferred by distance separation, presumably due to the absence of temporal-envelope differences.

Processed BRIRs: ‘ST’

- There is a significant increase in room-position responses compared with ‘SO’ conditions, confirming the existence of temporal-envelope effects. This is in contrast to ‘real-room’ conditions, indicating that ITD effects play a relatively minor role.

Conclusions

In dichotic conditions, room-position can dominate a same-sex talker difference. The binaural advantage seems primarily due to ILD.

Position can also compete with a talker difference in dichotic conditions, especially if there is a distance separation. Effects of distance seem due to temporal-envelope cues.

ITD-based effects seem to play a relatively minor role in selective listening in real-room reverberation.

References

5. Kidd, et al., 2005 Other aspects of position in a room may play a part however. For instance, the decrease in Direct-to-Reverberant energy Ratio (DRR) with increasing distance between listener and source is a cue to perceived distance (Zahorik, 2002). More generally, reflections distort sounds (spectral- and temporal-envelopes), and this varies with position, potentially giving rise to position-specific ‘grouping’ cues.

Speaker

dummy head

-um

room

-um

dummy head in ears

-um

BRIR

-um

monaural speech

-um

recording

-um

spatial

-um

room

-um

listening