Hearing Research 323 (2015) 32–39

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Use of binaural and monaural cues to identify the lateral position of a virtual object using echoes☆

Daniel Rowan*, Timos Papadopoulos [1], David Edwards [2], Robert Allen

Institute of Sound and Vibration Research, University of Southampton, UK

ARTICLE INFO

Article history:
Received 12 December 2014
Received in revised form
23 January 2015
Accepted 27 January 2015
Available online 7 February 2015

ABSTRACT

Under certain conditions, sighted and blind humans can use echoes to discern characteristics of otherwise silent objects. Previous research concluded that robust horizontal-plane object localisation ability, without using head movement, depends on information above 2 kHz. While a strong interaural level difference (ILD) cue is available, it was not clear if listeners were using that or the monaural level cue that necessarily accompanies ILD. In this experiment, 13 sighted and normal-hearing listeners were asked to identify the right-vs.-left position of an object in virtual auditory space. Sounds were manipulated to remove binaural cues (binaural vs. diotic presentation) and prevent the use of monaural level cues (using level roving). With low- (<2 kHz) and high- (>2 kHz) frequency bands of noise, performance with binaural presentation and level rove exceeded that expected from use of monaural level cues and that with diotic presentation. It is argued that a high-frequency binaural cue (most likely ILD), and not a monaural level cue, is crucial for robust object localisation without head movement.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Humans have the capacity for using echoes to determine certain features of their general environment and particular objects within it (Kolarik et al., 2014). This capability is functionally important to some blind people in their daily life for spatial awareness and navigation (Thaler, 2013). Understanding the auditory cues involved is important for the development of effective technology and training to enhance echolocation ability, and for understanding the effects of hearing impairment and hearing technology on it. Echolocation might be useful for a variety of tasks, such as the detection and localisation of objects, the detection and tracking of edges, and the discrimination of the material and density of objects. The current paper is focused on the use of echoes to localise objects.

In a previous paper, we distinguished between two general strategies for object localisation: 'scanning' (object detection with relative motion between source/receiver and object) and 'searching' (using information within the echo) (Rowan et al., 2013). A series of papers have investigated the searching strategy, showing that many sighted, visually impaired and blind people are able to accurately identify the right-left position of a real or virtual object under certain conditions (Després et al., 2005; Dufour et al., 2005; Rowan et al., 2013) and can discriminate changes in azimuthal position of one object relative to another (Teng and Whitney, 2011; Teng et al., 2011). Some blind people, at least, can also accurately judge object azimuthal location (Rice, 1967; Thaler et al., 2011). One might expect that azimuthal localisation of an object depends on binaural cues in a similar manner to that expected from sound-source localisation. We previously reported acoustical analyses showing azimuth-dependent interaural level differences (ILDs) and interaural time differences (ITDs) with a medium-density fibreboard (MDF) board at different distances and orientations (Papadopoulos et al., 2011). Interaural level differences were present above 2 kHz for all orientations investigated. More subtle ILDs were present below 2 kHz and ITDs in the waveform fine-structure were also present below 2 kHz only when there were specular reflection paths (i.e. involving the face

Abbreviations: BBN, Broadband noise; HPN, High-pass noise; ILD, Interaural level difference; IR, Impulse response; ITD, Interaural time difference; KEMAR, Knowles Electronic Manikin for Acoustics Research; LPN, Low-pass noise; MDF, Medium-density fibreboard

☆ Portions of this work were presented at the British Society of Audiology Short papers meeting on Experimental Studies of Hearing and Deafness, Southampton, UK, September 2009.

* Corresponding author. Institute of Sound & Vibration Research, University of Southampton, Southampton, Hants, SO17 1BJ, UK. Tel.: +44 (0)23 8059 2928, +44 (0)23 8059 2288; fax: +44 (0)23 8059 3190.

E-mail addresses: dr@isvr.soton.ac.uk, d.rowan@southampton.ac.uk (D. Rowan).

[1] Biodiversity Institute, Department of Zoology, University of Oxford; Machine Learning Research Group, Department of Engineering Science, University of Oxford.

[2] Yeovil District Hospital NHS Foundation Trust, Yeovil.

http://dx.doi.org/10.1016/j.heares.2015.01.012
0378-5955/© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

of the board) to both ears. (Humans are insensitive to ITD in the waveform fine-structure above approximately 1.5 kHz, Brughera et al., 2013.) Subsequent listening experiments revealed that information above 2 kHz was crucial for robust localisation of objects with a variety of orientations (Rowan et al., 2013). It is therefore tempting to assume that ILDs provided the information above 2 kHz. However, a change in ILD necessarily produces a change in level at one or other ear. While listeners seemed unable to use the monaural level cue for discriminating ILD with pure-tones associated with fused auditory images (Bernstein, 2004), the complex stimuli used in our previous object localisation experiments seemed to produce diffuse, complex images (Rowan et al., 2013), perhaps because of the complex frequency dependence of the ITDs and ILDs (Papadopoulos et al., 2011). It is unclear if performance on these object localisation experiments with stimuli above 2 kHz was due to ILD, and thus a binaural cue, or the associated monaural level cue. The current study addressed this issue using a virtual object localisation task (Rowan et al., 2013) and by limiting the access to binaural cues and monaural level cues.

2. General methods

Approval of the Faculty Ethics Committee was obtained before commencing the experiment. Thirteen otologically and ophthalmologically normal (excluding corrected short-sightedness) listeners (four female; 22–37 years old with mean of 25.5 years) were recruited from the university population; most were postgraduate students of the Institute of Sound & Vibration Research. They had hearing threshold levels below 20 dB HL for frequencies from 0.25 to 8 kHz at octave intervals with interaural asymmetries of no more than 10 dB at a single frequency. All but one were experienced with psychoacoustical experiments and four had previously participated in similar echo-related experiments. Listeners were required to perform a single-interval recognition task (MacMillan and Creelman, 2005) implemented in custom-written MATLAB code. They heard a single bilateral stimulus simulating the board positioned to the right or left and were required to select one of two buttons accordingly. The 'correct' answer was displayed on a screen for 400 ms after a response was made. In each session, 90 trials were obtained for each stimulus condition, split across three blocks and in an order that was balanced across listeners.

Auditory virtual objects were created as follows. Binaural impulse responses (IRs), as used in Rowan et al. (2013), were obtained between the electrical input to a loudspeaker and the electrical outputs of in-ear microphones of KEMAR placed in an anechoic chamber containing the target object. The driver of the loudspeaker was positioned 0.25 m below and 0.05 m in front of KEMAR's interaural axis, which itself was 0.975 m above the chamber's grid floor. The object and its orientations were as used in Rowan et al. (2013): a 0.55-m × 0.55-m (0.01 m deep) medium-density fibreboard (MDF) board placed vertically with its centre at the same height as KEMAR's interaural axis and at either 17° azimuth to the right or left. (Note that this is the same sized board as we used in both our previous papers; we incorrectly reported this as 0.55 m^2 in Rowan et al., 2013). The board was either parallel to the interaural axis ('flat'; specular reflection path to one ear or neither ear) or tilted so that the two vertical edges of the board were equidistant from the centre of KEMAR's head ('angled'; specular reflection path to both ears). The IRs were convolved with digitally synthesised bands of filtered Gaussian noise to produce binaural stimuli presented over Etymotic Research ER2 insert earphones to listeners seated upright in an audiometric booth. Digital signals were played out at a sampling rate of 44.1 kHz and with 16-bit amplitude resolution using a Creative Extigy soundcard. Each band of noise was generated independently prior to convolution. Approximately 1-s-long bands of noise were convolved with the IRs; the resultant stimuli were then windowed about the centre of the noise to achieve a duration of 400 ms, using 40-ms-long raised-cosine onset/offset ramps. A duration of 400 ms ensured high scores in the baseline board and bandwidth conditions used (Rowan et al., 2013), which was necessary in order to provide the potential for listeners to exceed the criteria for guessing and use of cues other than the monaural level cue. The ramps and the focus on the long temporal component where the emission and echo interfere (see Rowan et al., 2013; for a fuller discussion of these components), in combination with the filtering, were used to avoid energy below 2 kHz in the high-pass stimulus. A pilot study with 18 listeners found that performance with unwindowed (as in Rowan et al. (2013)) and windowed stimuli (as in the current study) was similar for 80–400-ms-long noise bands. As in our previous paper, the rms levels of pre-convolution noise stimuli were adjusted to produce the desired level (dBA) in a IEC 711 coupler when convolved with IRs with the board at 0.9 m and 0°.

Three stimulus manipulations were used in this experiment. Firstly, to investigate the use of cues at low and high frequencies, low-pass noise (LPN) from 20 Hz to 2 kHz and high-pass noise (HPN) from 2 kHz to 20 kHz was used. Broadband noise (BBN) from 20 Hz to 20 kHz was also used in the first session. However, the sensitivity of the insert earphones reduces steeply above approximately 12 kHz producing an upper frequency limit of approximately 12 kHz in practice (as was also the case with Rowan et al. (2013)). Filtering was achieved by digitally setting the amplitude of the spectrum outside the pass-band to zero in order to produce extremely steep slopes; this occurred before convolution. Secondly, to disrupt the use of a monaural level cue, the presentation level of a stimulus on a trial varied randomly (i.e. was 'roved') using a rectangular, uniform distribution extending from 50 to 80 dBA. The uppermost score that could be achieved with a monaural level cue was determined using a statistical model (Dai and Kidd, 2009) based on the output of a cochlear model (Chen et al., 2011). The cochlear model was used to estimate the magnitude of the 'unwanted' monaural level cue and level rove, both in terms of excitation level at the output of the cochlea. The former was calculated for stimuli presented at 65 dBA (calibrated as in the experiment, in principle), the mean level used in the experiment. This is plotted for the flat and angled orientations in Fig. 1 (a) and (d), respectively. The excitation level rove was calculated by comparing the excitation levels produced at the two extreme levels (50 and 80 dBA) for the two object positions. This is plotted for the flat and angled orientations in Fig. 1 (b) and (e), in both cases with the board to the right (similar results were found with the board to the left). The maximum score expected from use of the monaural excitation level cue given the excitation level rove at the output of each auditory filter using the Dai & Kidd model is plotted in Fig. 1 (c) and (f). This was then repeated at the lowest and highest stimulus presentation levels. The monaural excitation level cue was slightly larger when the stimulus was presented at 50 dBA (and slightly smaller at 80 dBA). The peak percent correct within the relevant stimulus bandwidth from the data at 50 dBA was taken as the maximum score possible expected from the monaural level cue: 62% for the LPN and angled orientation, 81% for HPN and angled orientation, and 66% for the HPN and flat orientation. In order for a listener's performance to be statistically significantly higher than these values with 99% confidence, given 90 trials per condition, they must score ≥75% for the LPN and angled orientation, ≥92% for HPN and angled orientation, and ≥79% for the HPN and flat orientation.

The third stimulus manipulation was used to investigate performance expected from monaural cues only. Stimuli were either presented in the standard, 'binaural' condition (as in Rowan et al. (2013)) or the sound for the right ear was presented to both ears in the 'diotic' condition. It is assumed that the diotic condition

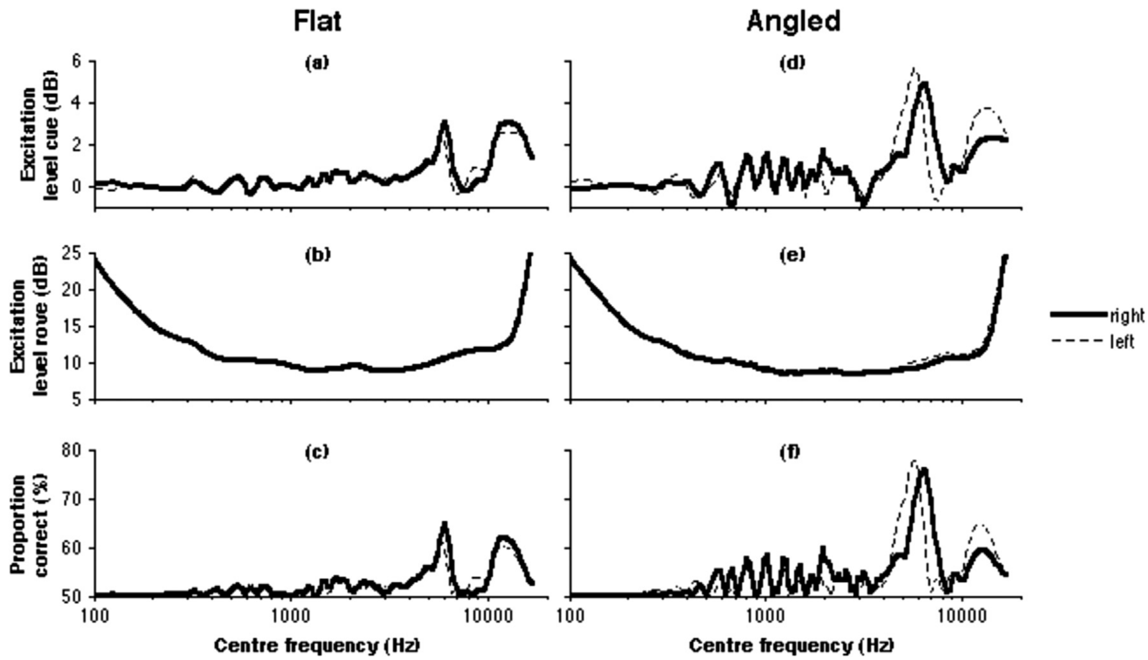

Fig. 1. The left and right columns refer to the 'flat' and 'angled' board orientations, respectively. The magnitude of the 'unwanted' monaural excitation level cue at the output of a cochlear model (Chen et al., 2011) for the stimuli used in the current experiment is plotted in the top row. The excitation level rove at the output of the cochlear model corresponding to an input level rove of 30 dB is plotted in the middle row. The maximum score expected from use of the monaural excitation level cue given the excitation level rove based on a statistical model (Dai and Kidd, 2009) is plotted in the bottom row.

contains the same monaural cues, including the monaural level cue, that are present in the binaural condition but without the binaural cues and that any bilateral advantage in using the monaural cues in the 'binaural' condition is also present in the 'diotic' condition.

Session 1 was organised in similar way to that in Experiment 2 from our previous paper (Rowan et al., 2013). Following brief familiarisation with the task, each listener completed 90 trials at five distances (0.6–1.8 m) with BBN and flat orientation. Sessions 2 and 3 contained the main experimental conditions and were identical: 90 trials were collected from 12 conditions built from the three factors *stimulus/orientation* (three levels: LPN with angled orientation, HPN with angled orientation and HPN with flat orientation), *presentation* (two levels: binaural vs. diotic) and *level roving* (two levels: rove or no rove). A distance of 0.9 m was used, following previous research (Després et al., 2005; Dufour et al., 2005; Rowan et al., 2013). In total, 180 trials were collected in our main conditions.

The data were analysed using signal detection theory to remove response bias. The statistic *d'* was calculated from raw frequencies of responses for each listener (MacMillan and Creelman, 2005) and then transformed back to (unbiased) overall percent correct for the purposes of plotting and parametric analysis, as in Rowan et al. (2013). These transformed percent correct scores are presented as box plots; the grey area represents the 99% range expected from guessing given the number of trials per condition. Statistical analysis (using SPSS v. 20) was conducted on arcsine-transformed scores, using parametric methods when the data were at least approximately normally distributed and paired samples, two-tailed tests.

3. Results

3.1. Session 1: distance with flat orientation

Fig. 2 plots the results from Session 1 of the current experiment alongside those from Session 1 of Experiment 2 from Rowan et al. (2013). There is a clear trend for listeners in the current experiment to achieve higher scores than those in the previous experiment, up to 1.5 m; this difference was statistically significant at 0.6 m only (independent samples *t*-test: $t_{24} = 3.5$, $p = 0.002$; $p > 0.1$ at other distances) and will be discussed in Section 4.1. Both experiments agree that performance worsens with increasing

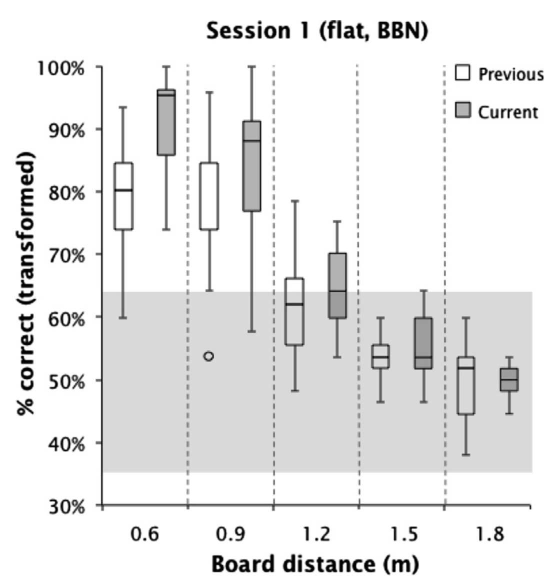

Fig. 2. Summary of results from Session 1 of the current experiment (n = 13) compared to those from Session 1 of Experiment 2 from Rowan et al. (2013) (n = 13), presented as box plots. Scores are expressed in terms of transformed percent correct of right-vs.-left judgements (see final paragraph of Section 2). Circles indicate values for individual listeners if greater than 1.5 times the inter-quartile range away from the nearest quartile; the grey area represents the 99% range expected from guessing. BBN: broadband noise.

distance, with the performance of all listeners being indistinguishable from chance by 1.5 m.

3.2. Sessions 2 & 3: monaural verses binaural cues

Fig. 3 presents the results from Sessions 2 and 3; recall that a board distance of only 0.9 m was used in these. In the rove conditions, the thin horizontal lines represent the maximum score expected from use of a monaural level cue and the thicker lines above them represent the scores individual listeners must reach or exceed to have performed statistically better than that, with 99% confidence.

The left column of Fig. 3 shows the results with the binaural presentation. The medians for two conditions are indistinguishable from the upper quartile. Overall and for most individual listeners, performance is similar in the two sessions (i.e. within 5%) and so the data were pooled across sessions for further statistical analysis. All listeners but one performed statistically better than chance in all conditions on both sessions. The results with no level rove reveal trends also apparent in our previous study (Rowan et al., 2013). Firstly, performance with the flat orientation is statistically significantly better with the HPN (Fig. 3, left panel) than with the BBN (Fig. 2 at 0.9 m) ($t = 4.5$, $p = 0.001$). In the current experiment this could conceivably have arisen from a learning effect, although not in the previous experiments. Secondly, performance with the HPN is statistically significantly better with the angled than with the flat orientation (Wilcoxon $Z = 3.1$, $p = 0.002$). Thirdly, performance with the angled orientation is better with the HPN than with the

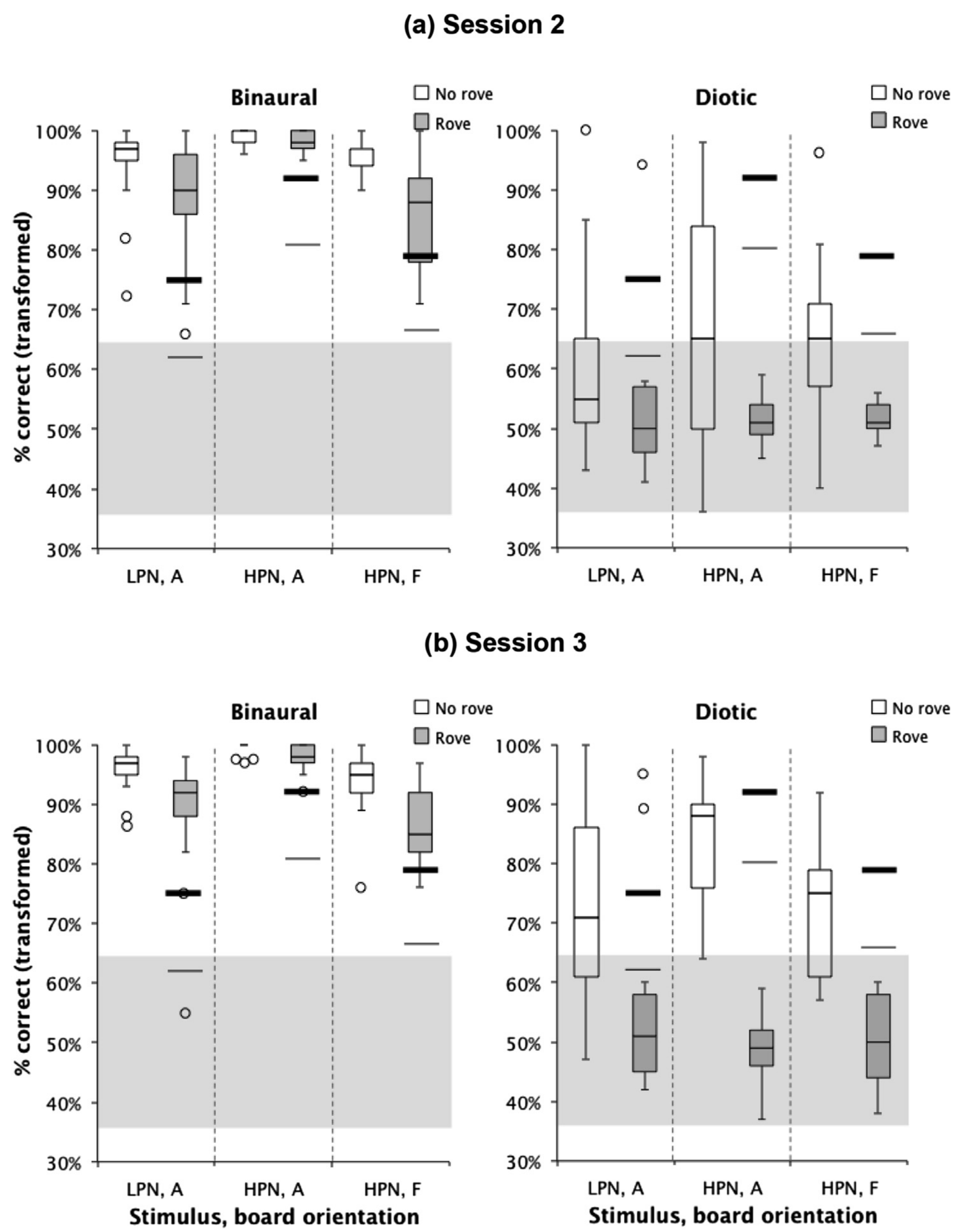

Fig. 3. The results from Sessions 2 (upper row) and 3 (lower row) of the current experiment using a board distance of 0.9 m, presented as in Fig. 2. Note that Session 3 was a replication of Session 2. In the rove conditions, the thin horizontal lines represent the maximum score expected from use of a monaural level cue and the thicker lines above them represent the scores individual listeners must reach or exceed to have performed statistically better than that, with 99% confidence. LPN, A: low-pass noise with the angled board orientation; HPN, A: high-pass noise with the angled orientation; HPN, F: high-pass noise with the flat orientation.

LPN (Wilcoxon $Z = 2.8$, $p = 0.005$). In the current experiment, adding the level rove tended to reduce performance (for LPN $t = 6.1$, $p < 0.001$; ceiling effect for HPN angled; for HPN flat $t = 1.6$, $p < 0.001$) although scores were statistically better than expected for use of a monaural level cue for most listeners.

The right column of Fig. 3 plots the results from the diotic presentation where listeners had access to monaural but not binaural cues (albeit presented to both ears). For the no level rove conditions there is a clear trend for performance to improve between sessions with several listeners improving by over 25%, albeit not or only marginally statistically significant: the median improvement was 6% ($t = 2.6$, $p = 0.03$), 13% ($t = 2.4$, $p = 0.03$) and 4% ($t = 1.8$, $p = 0.1$) in the LPN angled, HPN angled and HPN flat conditions, respectively. Overall and for most listeners, performance was considerably poorer with diotic compared to binaural presentation with no rove in both sessions ($t = 5.2–7.6$; $p < 0.001$). In Session 3, performance with no level rove was better with the HPN angled condition compared to both LPN angled ($t = 2.4$, $p = 0.04$) and HPN flat ($t = 5.4$, $p < 0.001$) conditions; performance with LPN angled and HPN flat conditions was similar ($t = 0.1$, $p > 0.1$).

Adding level rove to the diotic presentation reduced performance to chance levels for all listeners, all conditions and both sessions with no evidence of improvement between sessions. Exceptions were Listeners 6 and Listener 8 who scored substantially and statistically above that expected for use of a monaural level cue for the LPN angled level-rove condition in Session 3 or both Sessions 2 and 3, respectively. These two listeners were among the best performers throughout the experiment and were invited back for supplementary testing. Both reported being active musicians with music degrees (none of the other listeners had degree-level music qualifications), and using pitch to identify board position in the LPN condition. Listener 8 consistently scored high in a range of conditions featuring the LPN and level rove including binaural (as in left column of Fig. 3; 97%), diotic with both ears receiving information from the right (as in right column of Fig. 3; 84%) and left (96%) ear in the binaural conditions, monaural right (86%) and monaural left (75%). Listener 6 did not consistently score above chance with diotic or monaural presentation suggesting either that this listener could not use a non-level monaural cue consistently or that the seemingly above-chance scores were false positives.

4. Discussion

4.1. Overview

The results of the current experiment confirmed those of previous experiments in that (1) sighted people are able to identify the lateral position of a reflective surface at short range with binaural presentation of echolocation-related sounds (Dufour et al., 2005; Rowan et al., 2013) and (2) that ability is more accurate when the surface is angled (providing specular reflection paths to both ears) rather than flat and when the emission contains energy between 2 and 12 kHz rather than 0.1–2 kHz (Rowan et al., 2013).

Overall, the scores from the current sample of 13 listeners were typically higher than those from the sample of 13 listeners in Experiment 2 of Rowan et al. (2013). One possible explanation is that the stimuli in the current, but not the previous, experiment were manipulated to remove the emission-only and echo-only temporal components. Any adverse effects of the initial, emission-only component, e.g. via the precedence effect, would not have occurred in the current study. However, pilot testing showed similar scores on both types of stimuli with stimulus durations down to 80 ms. It is possible that the presence of the emission-only component is more important for shorter durations (e.g. as used in Després et al. (2005), Dufour et al. (2005), Rowan et al. (2013)). An alternative possible explanation for the typically higher scores in the current experiment is that listeners were recruited from a more select population (post-graduate audiology or engineering students); while the significance of any differences in intellectual abilities is questionable (Kidd et al., 2007) they might have been more familiar with listening experiments generally.

Interestingly, the magnitude of the differences in scores between the previous and current studies with sighted listeners is similar to the differences reported between 20 sighted and 12 blind listeners (Dufour et al., 2005) and between 15 sighted and 15 myopic listeners (Després et al., 2005). While the two groups in those two studies were similar in age, it is unclear if they were matched for other factors unrelated to visual status that might influence performance. Further research should document the magnitude of true inter-individual differences between people (estimated to be approximately 60% of the total variation for the BBN at 0.9 m in Experiment 2 of Rowan et al. (2013)) and identify factors involved in those inter-individual differences in order to enable further studies to provide more carefully controlled comparisons between populations (e.g. sighted vs. blind).

4.2. Monaural or binaural cues?

We suggested in our previous papers that the correct identification of right-left object position was due to the use of binaural cues, particularly ILD (Papadopoulos et al., 2011; Rowan et al., 2013). However, the use of a monaural cue, such as the monaural level cue that necessarily accompanies ILD, could not be ruled out. In the current study, level roving was applied during binaural presentation with the aim of selectively removing the monaural level cue. Similarly high scores were found with and without level roving. Those scores were sufficiently high with level roving to rule out the use of a monaural level cue in that roved condition. This indicates that a monaural level cue is not necessary to achieve or explain high scores without the rove. Diotic presentation of information at the right ear in the binaural condition was then used to investigate the expected performance using monaural cues alone, again with or without level roving and thus with and without the monaural *level* cue, respectively. Most listeners achieved scores above chance with diotic presentation, indicating the availability of monaural cues. Adding level rove reduced most (LPN) or all (HPN) scores to chance, indicating that level changes was the monaural cue with diotic presentation for most listeners. However, most listeners performed considerably worse with diotic, compared to binaural, presentation, indicating that monaural cues cannot explain performance during binaural presentation.

Adding the level rove during binaural presentation did reduce scores in both sessions. One potential explanation for this is that the monaural level cue was contributing to performance along with binaural cues in the absence of the level rove. Another potential explanation is that the level rove might have lead to a change in the ILDs at the output of the cochlea, although this was expected to be minimal based on the output of the model used to select the magnitude of the level rove (see Section 2). Alternatively, the rove might have reduced scores because it was distracting for listeners; the provision of more extensive experience with the level rove may have prevented this. Either way, we conclude that performance during the binaural presentation was primarily or entirely due to the use of binaural cues.

4.3. Which binaural cues?

Papadopoulos et al. (2011) presented an analysis of ITD and ILD cues arising from board geometries as used in the current and previous studies of object localisation (Després et al., 2005; Dufour et al., 2005; Rowan et al., 2013), including distances from 0.6 m to 3.0 m Fig. 4 plots similar analyses of ITD and ILD for convenience and to show more clearly perceptually relevant differences between board positions as a function of frequency. The differences (Δ) in interaural coherence (upper row), ITD (middle row) and ILD (lower row) for right minus left board positions for the flat (left column) and angled (right column) orientations for boards at 0.9 m are plotted. Binaural cues were calculated from the output of a gammatone filter bank in response to 5-s-long BBN of 0.1–12 kHz convolved with the relevant IRs. Interaural coherence and ITD were determined as the value and time delay, respectively, of the peak of the cross-correlation function of Hilbert envelope of the output of each filter; ILD was based on the ratio of the root-mean-square voltage/pressure at the output of each filter. Positive values were given to ITD and ILD when arriving earlier or having a higher pressure, respectively, in right ear. The BBN was also modulated at 125 Hz using a half-wave rectified tone prior to convolution with the IRs (Rowan and Lutman, 2007) in order to clarify the potential interaural coherence and envelope ITD cues available in general.

Interaural coherence for both positions and orientations was close to 1 across the entire frequency range, with differences in coherence between positions of 0.1–0.3 at the outputs of some filters above approximately 4 kHz. Presumably this is an indication that there are multiple reflection paths from board to ears and that these are not entirely interaurally symmetrical for the two board positions. While humans can detect a reduction in interaural coherence of as little as 0.02 (e.g. Gabriel and Colburn, 1981), it is unclear if this is actually a viable cue when (a) coherence reduces in such a narrow and high-frequency range as here and (b) the task is a single-interval (rather than 2/3AFC) one and hence the opportunity for direct comparison does not exist.

Differences in ITD were large below 2 kHz especially for the angled orientation although varied with frequency in a complex manner (Papadopoulos et al., 2011); that complex frequency dependence of ITD (and ILD) largely reflects acoustic interference between emission and echo (and the object being an imperfect reflector). The utility of this complex ITD information with the LFN is unclear. With the HPN, the bandwidth, the slopes of the filter and the steps taken to avoid spectral splatter (including relatively long onset/offset ramps) should have prevented the presentation of information within the frequency range of sensitivity to ongoing, fine-structure ITD (Brughera et al., 2013). Envelope-based ITD above 2 kHz was approximately zero for the flat orientation and 100–200 μs for the angled orientation. Recall that the actual stimuli presented to listeners in the present study had onset- and offset-ITD information removed. The higher scores typically observed in the current study (without onset ITD) compared to the previous (with onset ITD) indicate that onset ITD is not necessary for good performance, at least with relatively long-duration emissions; onset ITDs might be useful for abrupt emissions. While the emissions actually presented to listeners were not purposely modulated, salient envelope fluctuations at the output of cochlear band-pass filtering might occur. The perceived spatial position of high-frequency 'unmodulated' noise can influenced with envelope-based ITD (e.g. Trahiotis and Bernstein, 1986) although is dominated by ILD when envelope-based ITD and ILD are presented in combination during sound-source localisation (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002).

Assuming that differences in ILD of approximately 1 dB and higher are discriminable (e.g. Hartmann and Constan, 2002), perceptually relevant differences in ILD between board positions are apparent above approximately 1 kHz with flat orientation and above approximately 500 Hz with the angled orientation. An ILD cue is therefore available for the LPN with both orientations; this is apparent in our experiment by the availability of a monaural level cue. The relative contribution of ILD vs. ITD below 2 kHz with the angled orientation is unclear. Above 2 kHz, ILD provides a robust cue for both orientations.

It seems mostly likely that performance with the HPN in the current and previous studies (Rowan et al., 2013) was primarily due to the use of ILD, although further research is required to determine whether other binaural cues, such as the reduction in interaural coherence, make a contribution. Onset ITD might be important for other stimulus configurations. For BBN as used by Dufour et al. (2005) and Després et al. (2005), it seems that high-frequency ILD and low-frequency fine-structure ITD cues are combined even though this seems to be at the expense of object localisation scores (Rowan et al., 2013).

4.4. What is the low-frequency non-level monaural cue?

One listener demonstrated consistent ability to use a non-level monaural cue with the low-pass band of noise. According to the listener, a change in pitch was used. This might relate to 'repetition pitch', which arises when a noise signal is added to a delayed copy of itself (i.e. ripple noise with one iteration, e.g. Yost et al., 1978). For the geometry considered here (with a source 0.25-m below and 0.05-m in front of centre of head and reflector at 0.9-m distance and 17° azimuth), the repetition pitch corresponds to approximately 200 Hz. The change in repetition pitch for diotic presentation

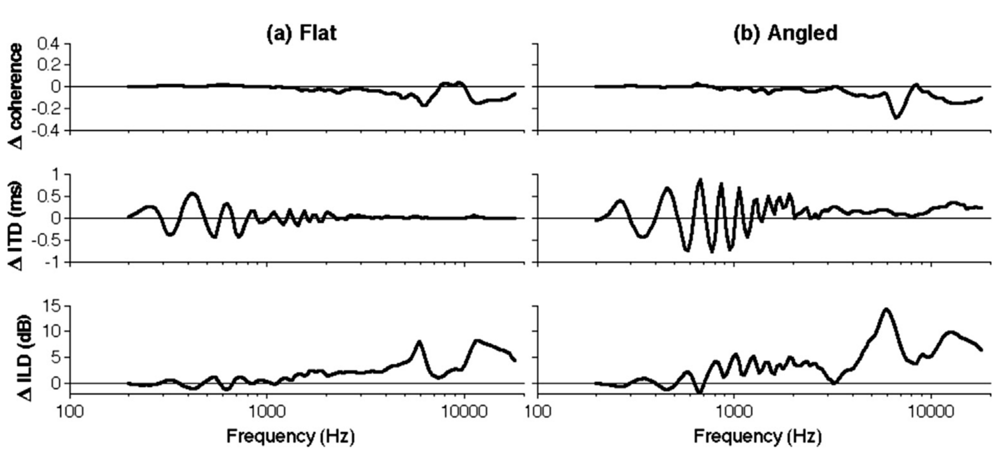

Fig. 4. For a board distance of 0.9 m, the difference (Δ) in interaural coherence (upper row), ITD (middle row) and ILD (lower row) for right minus left board positions for the flat (left column) and angled (right column) orientations. See main text for details.

between the board right and left positions is related to the ITD arising from the board at 17° azimuth (approximately 150 μs, Papadopoulos et al., 2011): approximately 6 Hz or 3%. While this is change in frequency is just about discriminable with pure tones (e.g. Wier et al., 1977), it would be considerably more difficult for repetition pitch (Yost et al., 1978), especially when the change in repetition pitch is not available on a single trial as in our experiment. Whether this is the cue the listener used is unclear.

4.5. What are the implications for real-world human echolocation?

While the findings of the current and previous experiments (Després et al., 2005; Dufour et al., 2005; Teng and Whitney, 2011; Teng et al., 2011; Rowan et al., 2013) demonstrate that sighted and blind listeners can use acoustic cues to identify and discriminate object horizontal position, they do not demonstrate that listeners have a clear spatial percept that corresponded to what would be expected from sound-source localisation and might be required for successful object localisation using searching in daily life. In fact, there is reason to doubt that listeners did hear clear spatial percepts from the complexity of the interaural cues in the stimuli and from the reports of listeners. There is evidence from other studies that accurate judgement of object location using searching is possible (Rice, 1967; Thaler et al., 2011) and more research is required to investigate this further.

The sighted listeners in our studies had relatively little experience with object localisation compared to what one might imagine for a blind person regularly using echolocation. There is some evidence that people with long-term blindness can learn to use monaural cues more effectively than sighted listeners for horizontal sound-source localisation, and thus possibly for object localisation. For example, Doucet et al. (2005) found that some (but not all) blind but not sighted people have accurate horizontal sound-source localisation when forced to only use monaural cues by unilateral ear plugging. The significance of this outside of the laboratory is unclear.

One of the limitations of our experiments investigating the acoustic cues used for localising an object is that they all used a flat surface of the same size: 0.5 m × 0.5 m. The size of the surface, as well as it's orientation (Rowan et al., 2013), is expected to strongly influence the availability of low-frequency information due to the acoustical processes generating the echo. Ashmead and Wall (2002) have reported that information below 200 Hz can be used in echolocation for larger surfaces, such as walls. However, it is not clear if Ashmead & Wall's experimental task specifically relates to azimuthal object localisation using the searching strategy; we expect that the acoustic cues and auditory processes necessary for successful echolocation depend on the particular task.

The importance of information and ILDs above 2 kHz for robust object localisation using searching has bearing on the question of how hearing impairment and prostheses affect echolocation, which usually affect the ability to hear and accurately encode sounds above 2 kHz (Moore, 2007). The LPN conditions in our experiments therefore provide a crude and extreme model for the loss of audibility with hearing impairment: object localisation using searching can be strongly affected by high-frequency hearing loss. Further research is necessary to investigate the effects of less extreme impairments in hearing thresholds and the effects of impairments in supra-threshold abilities. Fig. 4 indicates that there is ILD information above 10 kHz; conventional audiometric testing (which tests up to 8 kHz) and amplification (which extends up to 4 kHz) might be inadequate to reveal and ameliorate, respectively, auditory deficits relevant to some blind people. The signal processing in hearing prostheses, even if providing sufficient audibility of acoustic echolocation cues, might hinder object localisation (Simon and Levitt, 2007) through the distortion of ILDs by amplitude compression (Wiggins and Seeber, 2011) and through echo cancellation.

Finally, as discussed previously (Rowan et al., 2013) and reviewed extensively by others (e.g. Kolarik et al., 2014; Stoffregan and Pittenger, 1995), echolocation is a rich, complex and diverse multi-sensory, motor and cognitive phenomenon. The numerous factors involved are not captured in our experiments to date. However, our experiments provide insight into the auditory potential of, and acoustic cues involved in, at least one aspect of echolocation.

5. Conclusions

(i) Sighted people can use acoustic cues to identify the right-vs-left position of a flat board with a searching strategy, confirming previous research.
(ii) Performance is best with binaural presentation, during which it is primarily or entirely due to binaural cues.
(iii) A high-frequency (above 2 kHz) binaural cue, most likely ILD, is important for robust object localisation using a searching strategy.
(iv) Some individuals can use a currently undetermined monaural low-frequency cue that is not related to changes in overall level, at least to distinguish between two board positions.

Acknowledgements

David Edwards was supported by an RCUK studentship through a Basic Technology Programme grant to the Bio-Inspired Acoustical Systems EP/C523776/1 project (www.biasweb.org). Box-plots were created using a (modified) template produced by Vertex42 (www.vertex42.com/ExcelTemplates/box-whisker-plot.html). The MATLAB code for the gammatone filter bank was produced by Dan Ellis and acquired from www.ee.columbia.edu/~dpwe/resources/matlab/gammatonegram. A database of our IRs covering a wide range of conditions, including those in the current experiment, can be downloaded via: www.southampton.ac.uk/engineering/research/projects/make_like_a_bat.page. Many thanks to the two reviewers and the Associate Editor, Brian Moore, for helpful comments during the review process.

References

Ashmead, D., Wall, R., 2002. Low frequency sound as a navigational tool for people with visual impairments. J. Low. Freq. Noise Vib. Act. Control 21, 199–205.
Bernstein, L.R., 2004. Sensitivity to interaural intensitive disparities: listeners' use of potential cues. J. Acoust. Soc. Am. 115, 3156–3160.
Brughera, A., Dunai, L., Hartmann, W.M., 2013. Human interaural time difference thresholds for sine tones: the high-frequency limit. J. Acoust. Soc. Am. 133, 2839–2855.
Chen, Z., Hu, G., Glasberg, B.R., Moore, B.C.J., 2011. A new method of calculating auditory excitation patterns and loudness for steady sounds. Hear. Res. 282, 204–215.
Dai, H., Kidd, G., 2009. Limiting unwanted cues via random rove applied to the yes-no and multiple-alternative forced choice paradigms. J. Acoust. Soc. Am. 126, EL62–EL67.
Després, O., Candas, V., Dufour, A., 2005. Auditory compensation in myopic humans: involvement of binaural, monaural, or echo cues? Brain Res. 1041, 56–65.
Doucet, M.E., Guillemot, J.P., Lassonde, M., Gagné, J.P., Leclerc, C., et al., 2005. Blind subjects process auditory spectral cues more efficiently than sighted individuals. Exp. Brain Res. 160, 194–202.
Dufour, A., Després, O., Candas, V., 2005. Enhanced sensitivity to echo cues in blind subjects. Exp. Brain Res. 165, 515–519.
Gabriel, K.J., Colburn, H.S., 1981. Interaural correlation discrimination: 1. Bandwidth and level dependence. J. Acoust. Soc. Am. 69, 1394–1401.
Hartmann, W.M., Constan, Z.A., 2002. Interaural level differences and the level-meter model. J. Acoust. Soc. Am. 112, 1037–1045.
Kidd, G.R., Watson, C.S., Gygi, B., 2007. Individual differences in auditory abilities. J. Acoust. Soc. Am. 122, 418–435.

Kolarik, A.J., Cirstea, S., Pardhan, S., Moore, B.C.J., 2014. A summary of research investigating echolocation abilities of blind and sighted humans. Hear. Res. 310, 60–68.
MacMillan, N.A., Creelman, C.D., 2005. Detection Theory: a User's Guide. Lawrence Erlbaum Associates, Inc, Mahwah, N.J.
Macpherson, E.A., Middlebrooks, J.C., 2002. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. J. Acoust. Soc. Am. 111, 2219–2236.
Moore, B.C.J., 2007. Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, second ed. John Wiley & Sons Ltd, Chichester.
Papadopoulos, T., Edwards, D.S., Rowan, D., Allen, R., 2011. Identification of auditory cues utilized in human echolocation — objective measurement results. Biomed. Signal Process. Control 6, 280–290.
Rice, C.E., 1967. Human echo perception. Science 115, 656–664 (80-.).
Rowan, D., Lutman, M.E., 2007. Learning to discriminate interaural time differences at low and high frequencies. Int. J. Audiol. 46, 585–594.
Rowan, D., Papadopoulos, T., Edwards, D., Holmes, H., Hollingdale, A., et al., 2013. Identification of the lateral position of a virtual object based on echoes by humans. Hear. Res. 300, 56–65.
Simon, H.J., Levitt, H., 2007. Effect of dual sensory loss on auditory localization: implications for intervention. Trends Amplif. 11, 259–272.
Stoffregan, T.A., Pittenger, J.B., 1995. Human echolocation as a basic form of perception and action. Ecol. Psychol. 7, 181–216.
Teng, S., Puri, A., Whitney, D., 2011. Ultrafine spatial acuity of blind expert human echolocators. Exp. Brain Res. 216, 483–488.
Teng, S., Whitney, D., 2011. The acuity of echolocation: spatial resolution in the sighted compared to expert performance. J. Vis. Impair. Blind 105, 20–32.
Thaler, L., 2013. Echolocation may have real-life advantages for blind people: an analysis of survey data. Front. Physiol. 4, 1–8.
Thaler, L., Arnott, S.R., Goodale, M.A., 2011. Neural correlates of natural human echolocation in early and late blind echolocation experts. PLoS One 6, e20162.
Trahiotis, C., Bernstein, L.R., 1986. Lateralization of bands of noise and sinusoidally amplitude-modulated tones: effects of spectral locus and bandwidth. J. Acoust. Soc. Am. 79, 1950–1957.
Wier, C.C., Jesteadt, W., Green, D.M., 1977. Frequency discrimination as a function of frequency and level. J. Acoust. Soc. Am. 61, 178–184.
Wiggins, I.M., Seeber, B.U., 2011. Dynamic-range compression affects the lateral position of sounds. J. Acoust. Soc. Am. 130, 3939–3953.
Wightman, F.L., Kistler, D.J., 1992. The dominant role of low-frequency interaural time differences in sound localization. J. Acoust. Soc. Am. 91, 1648–1661.
Yost, W.A., Hill, R., Perez-Falcon, T., 1978. Pitch and pitch discrimination of broadband signals with rippled power spectra. J. Acoust. Soc. Am. 63, 1166–1173.