

SUBJECTIVE EVALUATION OF AN EMERGING THEORY OF LOW-FREQUENCY SOUND SOURCE LOCALIZATION IN CLOSED ACOUSTIC SPACES

AJ Hill
MOJ Hawksford

Department of Engineering, University of Derby, Derby, UK
School of Computer Science & Electronic Engineering, University of Essex,
Colchester, UK

1 INTRODUCTION

The primary aim of sound reproduction is to accurately recreate naturally-occurring sound fields. Due to reproduction system limitations, however, compromises must be made to allow for practical sound reproduction. In regards to low-frequency sound reproduction, a common assumption is that humans cannot detect direction at low-frequencies or that in the presence of a broadband signal, low-frequency directional cues are not significant.

The research detailed in this paper takes an emerging theory of low-frequency sound source localization in closed acoustic spaces, developed by the authors over the past few years using objective analysis techniques^{1, 2}, and examines if it holds up when applied to subjective evaluation results. The theory, as it currently stands, states that virtual image acuity is strongly dependent on the inter-arrival time between a direct sound and its first reflection. This indicates that low-frequency localization is indeed possible, but depends on room topology, source placement and listener location in terms of the minimum localizable frequency.

Two distinct questions are considered in this work: one focuses on human perception of sound while the second looks at practical sound reproduction considerations. First, does the objectively-based theory agree with subjective evaluation results? This would give good support to the idea of low-frequency localization based on reflection times. Second, the relevance of these findings in terms of a broadband signal is discussed (i.e. even if low-frequencies are localizable, is this necessary in typical listening scenarios?).

The paper begins with a broad overview of existing knowledge on low-frequency sound source localization (Section 2), highlighting the conflicting viewpoints of many researchers. This is followed by a brief description of the current theory previously developed by the authors (Section 3). Section 4 is the core of this paper, detailing the subjective experimental method and examining the results of the investigation in terms of the existing theory and sound reproduction considerations. The paper is concluded in Section 5 with a number of suggestions for future work.

2 EXISTING KNOWLEDGE ON LOW-FREQUENCY SOUND SOURCE LOCALIZATION

Despite a considerable volume of previous research focusing on the issue of low-frequency localization, there still lacks a cohesive understanding of the subject. In many ways this is due to researchers using different experimental methods in an attempt to solve similar problems, but also as a result of key oversights in certain aspects of their work. This section first looks at the human hearing mechanism as it relates to localization and then concentrates on low-frequency localization in the context of sound reproduction.

2.1 Rayleigh's duplex theory

Sound localization in humans is generally most accurate across the horizontal plane (the plane covering front-to-back and left-to-right, parallel to the ground) which can be attributed to the evolutionary advantage of humans to locate threats, which tended to be ground-based.

Lord Rayleigh described the two key localization mechanisms in his pivotal work, the duplex theory³. Interaural time difference (ITD) occurs due to the physical separation of the ears. When a sound source is directly in front (or behind) a listener, the sound will reach the ears at precisely the same moment. As the sound source moves off-axis, however, the time arrival of the direct sound will differ by a fraction of a millisecond (the maximum delay is roughly 0.65 ms, but is dependent on the size of an individual's head). ITD can be approximately calculated using Eq. 2.1 (assuming the sound source is far from the head with only a direct path to one ear when off-center):

$$ITD = \frac{r(\varphi + \sin \varphi)}{c} \quad (2.1)$$

where the ITD is based on head radius (r), the source azimuth (φ) and the speed of sound in air, c .

This slight time difference is interpreted in the brain to correspond to a certain location on the horizontal plane. Although it is currently debated exactly what the mechanism is for this within the brain, the most commonly agreed explanation is set out in the Jeffress Model, whereby there is a set of neurons for each ear tied together in a delay line configuration⁴. Each neuron in this configuration is set to trigger only for a certain ITD between the ears, allowing for accurate localization without significant neural processing.

It is commonly said that the brain in fact operates on interaural phase difference (IPD) between the ears, rather than on ITD. This idea has been largely proven untrue though some extensive testing where the ITD and IPD of a free-field sound source was inspected over various azimuths. It was found that for a given azimuth the ITD changes very little with frequency while IPD changes quite considerably. IPD was therefore rejected as it would require significantly more complicated frequency-dependent localization mechanism within the brain^{5,6}.

Despite the lack of direct influence of IPD on localization, it is still thought to be a significant factor regarding the upper limit of ITD usefulness. As a sound is fed through the inner ear and into the brain, it is encoded as phase-locked discharges of the auditory nerve fibers (tuned to narrow frequency bands). Assuming the Jeffress model is correct, if there is more than one-half wavelength offset between the received signals at the ears, the neuron delay-line localization mechanism will have difficulty providing accurate localization information (although it is still operating based on ITD cues). This results in the upper limit of ITD accuracy, given by Eq. 2.2:

$$f_{\max} = \frac{c}{2r(\varphi + \sin \varphi)} \quad (2.2)$$

If an individual has a head radius (r) of 9 cm and a source azimuth of 90° off-axis (with a speed of sound, c , of 343 m/s), then the maximum frequency (f_{\max}) giving an accurate ITD cue is 741 Hz. If the sound source azimuth was perfectly on axis (0°) the upper limit is 1.91 kHz. Clearly, ITD accuracy is largely dependent on the azimuth of the sound source, resulting in a loss of localization precision at higher frequencies.

Interestingly, a number of researchers have suggested that ITD is indeed applicable at higher frequencies, as the hearing mechanism extracts ITD cues from the high-frequency energy envelopes, thus allowing for low enough phase differences to process for ITD cues^{6,7}. While an interesting point, it is beyond the scope of this work so will not be discussed further.

The second localization mechanism described by Rayleigh which makes up for ITD's issues at higher frequencies is interaural level difference (ILD). ILD is best understood as imagining a spotlight shining onto the head. When the light is directly in front (or behind) the head, both ears are illuminated equally. As the light moves off-axis, one ear ends up in the acoustic shadow of the head. This leads to a received level difference at each ear and provides the second localization mechanism of duplex theory.

In order for ILD to operate accurately, the head needs to provide a sufficient acoustic shadow. This leads to a necessary lower limit for ILD accuracy, as at frequencies with wavelengths sufficiently larger than an individual's head, there will be little to no acoustic shadow. This limit (f_{min}) is typically taken as the frequency at which the head diameter is less than one-third of the wavelength and is described by Eq. 2.3:

$$f_{min} \Big|_{\phi=\pi/2} = \frac{1}{3} \frac{c}{2r} \quad (2.3)$$

Taking the example given for ITD (an individual with a head radius of 9 cm), the minimum frequency for accurate ILD localization is 635 Hz. This assumes that the source is directly to either side of the head and this minimum frequency will increase as the source moves closer to center (thus decreasing the ILD). The minimum limit of ILD and the maximum limit of ITD indicate the complementary nature of the human localization mechanism according to Rayleigh's duplex theory.

In the context of low-frequency localization in closed spaces, there are a number of caveats to duplex theorem. Central to this is what has been termed the plausibility hypothesis⁸. In small rooms, early reflections cause an unnatural ITD. It is thought that humans are able to weight ITD cues based on their plausibility, whereby it is unconsciously understood what the maximum realistic ITD can be with a given head shape. Any received ITD cues falling outside of this range will be rejected and ILD will be used as the dominant localization cue (at least within the confines of duplex theory)⁸.

It may be argued that at low-frequencies if ITD has been rejected due to the implausibility of the cues, the remaining ILD cues will be insufficient (or non-existent) to allow for accurate localization in the closed space. This might lead to the conclusion that low-frequencies are not localizable in non-free fields. Fortunately, it is understood that ILD at low-frequencies can indeed be significant in small rooms due to strong early reflections. Previous research has found that ILD cues can aid localization for frequencies from 200 – 5000 Hz (200 Hz was the lower limit tested in this case)⁸.

Clearly, duplex theorem must account for a number of additional variables when in a reflective environment. No longer is there a clear distinction between the low-frequency centered ITD cues and the high-frequency focused ILD cues. This leads to support for the notion that low-frequencies can be localized in closed spaces.

2.2 The precedence effect

As discussed in the previous section, reflections have a significant influence on sound localization. While causing unreasonable ITD and usable ILD values at low-frequencies, the effect of reflections on sound localization also operates within the construct of what is known as the precedence effect. The precedent effect, first described by Haas⁹, becomes relevant when two or more sounds arrive at the ears at very short intervals (less than 30 ms apart). With such short time between sound arrivals, the first arriving sound is given highest weighting for localization purposes. Based on this phenomenon, it should be expected that any sound arriving within a 30 ms or so interval after the arrival of another sound should not have any substantial effect on localization. As will be revealed, however, this is not always the case.

It is important to note that this 30 ms integration (or fusion) interval is based on experiments conducted in an anechoic environment (with a single reflective surface) whereby the direct and reflected sounds were equal in level. The test signals in this case were human speech⁹. It has been

suggested that the precedence effect integration interval be extended to 50 ms for music signals and can in fact be up to (and beyond) 70 ms in some cases¹⁰. Clearly, the duration of the precedence effect is largely dependent on source signal content.

While the boundary between early and late arriving reflections is signal-dependent, extremely early arriving reflections must also be considered. Experiments have shown that reflections arriving approximately 10 ms or less after the direct sound cause the sound image to shift in perceived location towards the direction of the reflection^{10,11}. This effect is more exaggerated as the relative level of the reflection increases¹⁰. Reflections arriving beyond 10 ms often cause the sound image to broaden, which causes issues if attempting to pinpoint the source location.

In the context of this work, it is essential to understand what triggers the precedence effect, as when in small rooms there will be numerous small reflections which, according to the precedence effect, should not significantly affect the localization of a sound source. However, experiencing listening in small rooms suggests the possibility of a breakdown in the precedence effect at low-frequencies, causing confusion in localization.

This phenomenon was first alluded to in a suggestion of a possible trigger for the precedence effect being the onset rate of a sound (measured in Pa/s)¹². Essentially, the higher the onset rate the stronger the trigger for the precedence effect. It is important to note that onset duration is not the sole variable here. Onset amplitude must also be considered. For example, for a signal with an onset duration of 200 ms and an amplitude of 80 dB, the onset rate is 1.0 Pa/s. A signal with an onset duration and amplitude of 25 ms and 60dB, respectively, has an onset rate of 0.8 Pa/s. Although these two signals have greatly different onset durations, their onset rates are quite similar¹². This is thought to be essential in determining if the precedence effect will trigger or not.

This idea of a precedence effect trigger eventually led to the question of whether the precedence effect is even applicable to low-frequencies at all¹⁰. Low-frequencies, with their long periods, naturally have long onset durations. This causes low onset rates, which may prevent the precedence effect from fully taking hold for sound source localization. This is a reasonably acceptable explanation for the high level of confusion with low-frequency localization in small rooms. The precedence effect is not operating, causing all early reflections to smear the perceived location of a sound source. This assumption is central in the development of the theory outlined in the forthcoming sections of this work and provides support for its appropriateness.

2.3 Head-related transfer functions (HRTF)

During the discussion on the duplex theory, it became clear that there is a lack of distinction between front/back and above/below when relying solely on ITD and ILD localization cues. This is commonly referred to as the “cone of confusion”. This ambiguity is resolved with a number of methods. Slight head/body movements are commonly used to aid localization, particularly when challenged with localization in a reverberant environment¹³. Additionally, an individual's visual gaze has been shown to skew sound source localization in the direction of the visual cue¹⁴. This is commonly known as the “ventriloquist effect”.

The greatest aid in resolving ambiguities that stem from the duplex theory is each individual's head-related transfer function (HRTF). There has been a significant volume of previous research into how the physical structure of the human head and body play a role in localization. Primarily, the structure of the pinna (outer ear) provides short time reflections of an arriving sound. These reflections cause comb-filtering in the signal which the brain can interpret as characteristic frequency response notches of a particular arrival direction¹⁵.

Crucially to this work, is the point that HRTFs are reliant of broadband noises for proper localization. This is because the brain looks for a specific series of spectral notches to judge direction. For a spectrally sparse signal, this series will be incomplete, thus preventing accurate localization using this method⁷. As the experiments detailed in this paper use windowed sine and square waves, it needs to be noted that HRTFs are unlikely to aid in localization and the results need to be analysed

with this thought in mind. As HRTFs operate with very short time reflections, when considering low-frequencies this may not be a central issue, as ambiguities may exist in HRTFs in this range, even if dealing with a low-passed noise signal.

2.4 Sound reproduction considerations

As previously mentioned, a considerable volume of research has been conducted that focuses on whether multiple channel low-frequency sound reproduction is necessary in real-world listening applications. The conclusions stemming from these investigations largely disagree with one another, providing a clear separation of opinion (for most researchers): they either support the notion that low-frequency sound source localization is possible and/or important or they believe (through their work) that it is not possible and/or unimportant².

The majority of the previous research has been at least partially prompted by the fact that most surround sound configurations employ a single subwoofer for low-frequency content. Due to this practice, all ITD/ILD cues in the low-frequency band are eliminated. Additionally, the incorrect localization cues stemming from the single (often arbitrarily placed) subwoofer have the potential to detract from the broadband sound localization (due to the conflicting cues)⁷.

The thorough literature review previously conducted by the authors highlights some issues in past experiments on this topic². Nearly all of the work investigates a single room with a centrally-placed listening location and small number of source locations. The current project here has already stressed the fact that localization performance is dependent on room topology and source/listener location^{1,2}. If all these variables are not addressed, then it is impossible to develop a robust conclusion, hence the focus on developing a generalized theory in this work.

3 OBJECTIVE ANALYSIS OF LOW-FREQUENCY SOUND SOURCE LOCALIZATION

Research on this project to date has focused on developing an objectively-based generalized theory of low-frequency sound source localization^{1,2}. This was due to the issue stemming from previous work in this area solely utilizing subjective evaluations to support their theories. While subjective evaluation is essential in such a subject area, it needs to be supported by objective data.

The objective analysis was first based on a finite-difference time-domain (FDTD) acoustic model¹⁶, but was later switched to an image-source approach to avoid excessively long run-times. The simulated measurements were analyzed looking at ITD cues to determine the precise moment in time when these cues broke down to no longer indicate the correct direction of the source. This was performed for a series of measurement positions, measurement height, source locations and absorption levels.

Data analysis was conducted by windowing the simulated binaural room impulse responses (BRIR) by progressively wider windows (in 0.2 ms steps). From this, the complex frequency response was calculated for each window/BRIR combination. The time delay can then be calculated from the phase response using Eq. 3.1:

$$T(f) = -\frac{d(\text{phase}(f))}{d\omega} \quad (3.1)$$

where, $T(f)$ is the time delay (s) at frequency, f (Hz), $\text{phase}(f)$ is the phase (rad) at frequency, f , and $d\omega$ is the frequency bin spacing of the phase response. The time delay calculated in Eq. 3.1 allows for the ITD to be determined using Eq. 3.2:

$$ITD_R(f) = T_L(f) - T_R(f) \quad (3.2)$$

where, the received ITD, $ITD_R(f)$, is based on the left and right ear time delays, $T_L(f)$ and $T_R(f)$, respectively. Finally, the virtually perceived directional error, $\theta_{error}(f)$, can be found based on the difference between the result from Eq. 3.2 and the expected ITD for the given arrival angle, $ITD_{error}(f)$, with Eq. 3.3:

$$\theta_{error}(f) = \sin^{-1}\left(\frac{cITD_{error}(f)}{r}\right) \quad (3.3)$$

When the directional error exceeds a set threshold (usually around ± 5 -10 degrees is allowed) the uncorrupted localization time is known (the inter-arrival time between the direct sound and first reflection). Longer uncorrupted localization times should allow for accurate localization down to lower frequencies.

3.1 Previous findings

Upon inspection of the data gathered from the above-mentioned analysis procedure, it became clear that localization began to lose accuracy when the first reflection arrived at a measurement point. While a seemingly obvious conclusion, it is good that this could be confirmed by the thorough analysis. Expected uncorrupted localization times for two example scenarios are given in Fig. 3.1.

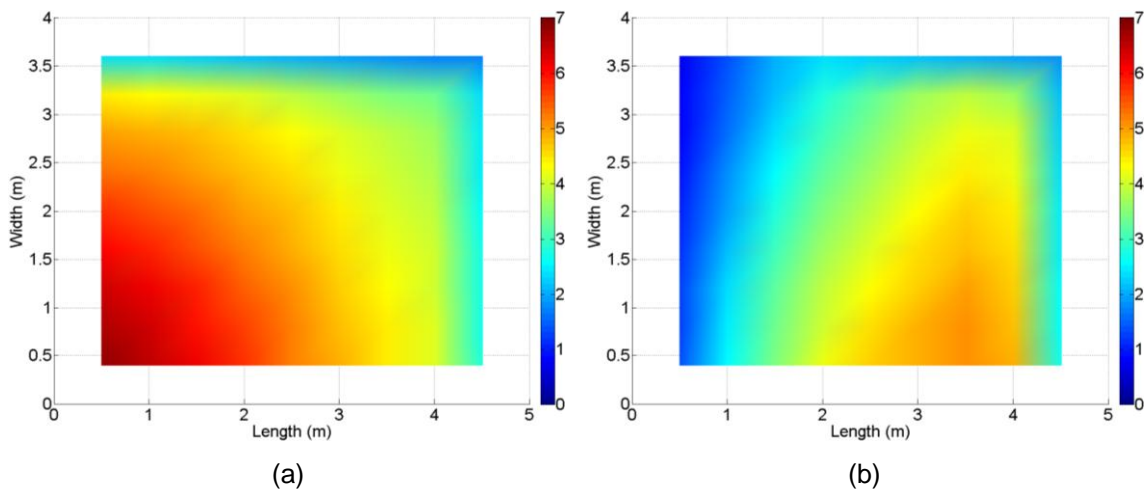


Fig. 3.1 Uncorrupted localization time (ms) (as determined by direct sound and first reflection arrival times) due to a subwoofer at (a) (0 m, 0 m) and (b) (1 m, 0 m) on the floor of a 5 m x 4 m x 3 m room (5% absorption) with a listener height of 1.8 m¹

An interesting observation from these results (and contrary to the initial hypothesis²) is that the closer a listener is located to a sound source does not imply better localization. This is evident in Fig. 3.1(b) whereby the source is one meter from the left wall, but listeners near the rear of the room along the left wall receive the most localization time while closer listeners along the wall have a much shorter time to locate the source. This finding clearly proves the importance of source and listener location on localization performance, which has not been a central focus of previous work on this subject.

Lastly, the average absorption coefficient (α) of the room was tested to determine if this had any bearing on the localization time (Fig. 3.2). The localization time with absorption taken into account, $T_L(\alpha)$, is described by Eq. 3.4¹.

It was found that absorption coefficients under 0.8 do not add a significant amount of localization time. Above 0.8, the additional localization time approaches infinity, indicative of an anechoic environment. The conclusion is, therefore, that absorption is only an important factor in outdoor situations such as music festivals or large-scale film screenings.

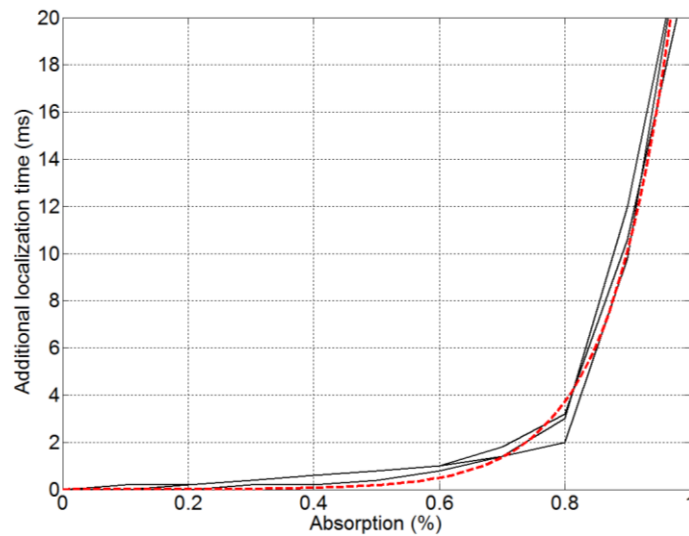


Fig. 3.2 Additional localization time versus absorption data (solid black lines) and best fit line (dashed red line) according to Eq. 3.4¹

$$T_L(\alpha) = T_L(0) + \frac{e^{10\alpha}}{800} \quad (3.4)$$

3.2 Re-examination of early reflections

Upon further review of previous research on localization it became clear that subjective evaluation results indicate that floor and ceiling reflections cause no corruption of azimuthal localization cues as they still arrive from the same direction on the horizontal plane⁸. This explains the slight disagreement between the calculations, theoretical results and real-world measurements for the work to date^{1,2}. Analysis of the subjective evaluation data from the current work will utilize the hypothesis that the uncorrupted localization time is dependent on the inter-arrival times between the direct sound and first “off-source” (non-floor/ceiling) reflection.

4 SUBJECTIVE ANALYSIS OF LOW-FREQUENCY SOUND SOURCE LOCALIZATION

The primary focus of the present investigation is to determine if the objectively-based theory of low-frequency sound source localization agrees with subjective evaluations. The objective theory gives a rough value of 1.4 uncorrupted wavelengths necessary for accurate localization. This is based on comparing findings to previous work and must be verified with sufficiently detailed and controlled subjective evaluations. The first in-depth listening test carried out to support and/or refine the theory is described herein.

4.1 Methodology

A blind listening test was designed in order to explore the localization theory. The test presented listeners with a randomized test signal: either a windowed sine-wave (tone-burst) or a windowed pseudo-square wave (odd-order harmonics up to the 15th harmonic). Each burst of the sine or square wave consisted of 100 cycles in order to avoid introducing additional spectral energy to the signal, but to also give listeners multiple opportunities to localize a sound (instead of a single continuous sound, which is well-known to be very difficult to accurately localize, due to the lack of onset to trigger the precedence effect^{8,12}). The test signal was repeated for 40 Hz, 65 Hz, 90 Hz and 115 Hz. In each instance, the sine or square wave was selected randomly.

In addition to the randomization of the test signal presented, the signal is sent to a random loudspeaker. The system consisted of four loudspeaker systems placed around a quarter of the perimeter of the space. Each system consisted of a high-pass filtered unit connected to a low-pass filtered subwoofer. All loudspeakers were hidden behind black drapes, hung from floor to ceiling.

Listeners were asked to indicate the perceived location of each test signal, by manual indication on their given data charts. They wrote a '1', '2', '3' or '4' at the perceived location, depending on which test signal was playing (4 signals were presented at each listening location). Listeners would repeat this process at each listening location (9 in total), giving 36 localization judgments for each listener.

Potential issues within the test room (10.6 m x 11.6 m x 9.1 m) were that a significant amount of large equipment (staging, truss, rigging equipment) was stored at the front 2 m of the space (near the top of the area in Fig. 4.1) and also that there was a small entrance corridor at the rear left (bottom left of Fig. 4.1). It is expected this caused results to be slightly skewed, especially for listening locations where the first reflection originated from the front wall or the rear left wall area.

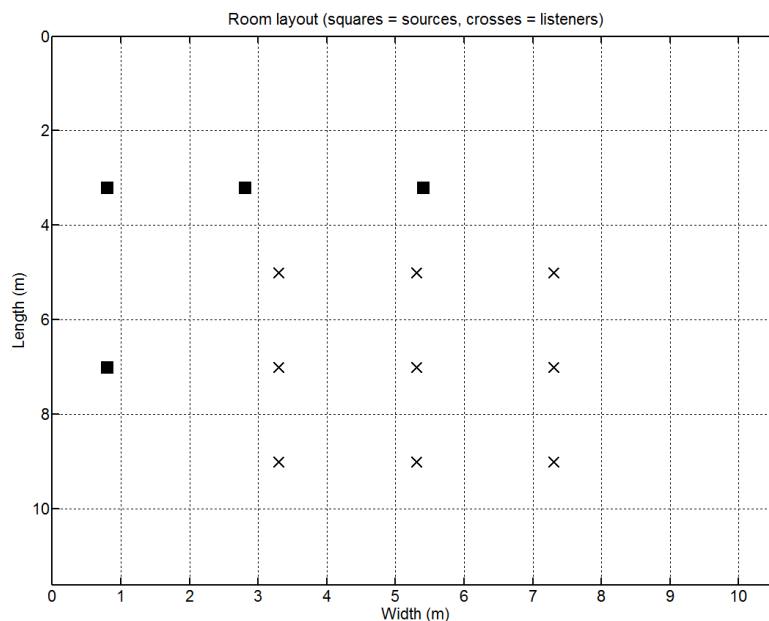


Fig. 4.1 Subjective evaluation source and listener location configuration (black drapes were suspended directly in front of the sources, traveling around the perimeter of the room)

Listening tests were carried out over two days in late February/early March 2014. Final year BSc (Hons) Sound, Light and Live Event Technology students from the University of Derby set up the system, which was fully tested and debugged before inviting listeners to participate.

Overall, there were 36 participants in the experiment, which were a mix of staff and students (all on audio engineering-related courses). Ages ranged from 18 to 55 and all participants were male except for one female student participant. The 36 participants resulted in a data set of 1296 localization judgments.

One issue that arose was that the loudspeakers had a very slight hiss, which was noticeable when seated in the front row of the listening grid. While this was obvious when test signals were not playing, it did not seem to effect listeners judging the direction of a sound, although could have given clues regarding the locations of the front sources. Another issue was if a test signal was stopped mid-burst, there would be a sharp click, due to the abrupt cut-off of the signal. This is not likely to have been a serious issue, since the signal was only stopped after all listeners had made their localization judgments.

4.2 Theory-based predictions

As previously mentioned, floor and ceiling reflections should be disregarded as they have the same azimuth as the direct sound, causing no horizontal localization confusion. This can be applied to the theory simply by extracting the relevant directional data from the image source model and outputting the time and direction data for each reflection. By comparing this data to the direct sound information, the first angle with a different azimuth can be pinpointed which better defines the time available for accurate localization in a space. The predicted uncorrupted localization times (based on the 1.4 wavelength requirement suggested by the theory) were converted to minimum localizable frequencies and are given in Fig. 4.2.

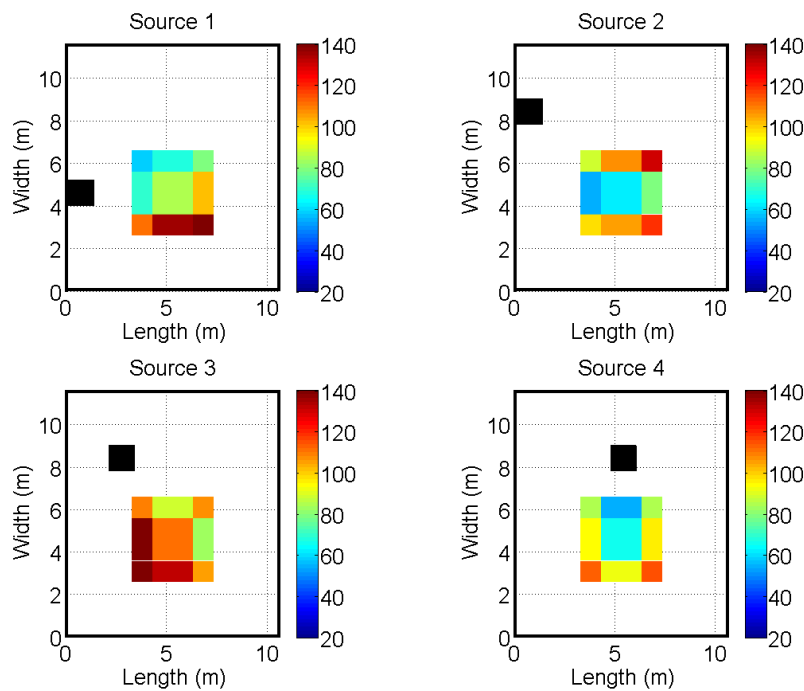


Fig. 4.2 Predicted minimum localizable frequencies (Hz) across the listening grid in the test space (black boxes indicate source location, layout corresponds to configuration shown in Fig. 4.1)

4.3 Results

Upon completion of all listening test trials, the data was entered into a MATLAB analysis program, alongside the automatically logged data from each trial (signal type (sine/square), output source, frequency). The true arrival angle for each listening location/loudspeaker combination was calculated to allow for the calculation of perceived localization error. It is important to note that due to the randomized nature of the experiments, there are not an equal number of occurrences of each source/signal/listener combination, so some combinations have one or two instances while others have closer to ten.

Plots were generated for each listening location and source combination (Figs. 4.3 – 4.6, located later in this paper). Large black stars indicate the true arrival angle for the source/listener combination. Small colored crosses indicate perceived source location for the windowed square wave signals while small circles indicate perceived source location for the windowed sine-wave signals. 40 Hz, 65 Hz, 90 Hz and 115 Hz data is indicated by the colors blue, cyan, yellow-green and red, respectively (which corresponds to the color scale used in the predicted localization frequency limits in Fig. 4.2).

Inspecting the results for source 1 (source located directly to the left of the listening grid) there are a few observations that can be made. Mostly noticeably is the widely varying localization data for the

top leftmost location. All perceived directions seem to be skewed towards the front of the room. A similar pattern can be detected for the other locations where the arrival angle is coming from behind the listener. At first glance, this may lead to the conclusion that there is a frontal bias in the localization mechanism, whereby backward-originating sounds are less accurately localized. Upon inspection of the experimental setup, however, a better explanation arises relating to the black drapes that were not extended across the rear of the room. Listeners, therefore, knew that there was no source located along the rear wall and it's likely this information resulted in the forward-bias in the results (Fig. 4.3).

The results for source 2 (source located near the corner of the listening grid, along the left side wall) show some encouraging agreement with the predictions (Figs. 4.2 and 4.4, respectively). The front and rear right locations were predicted to have the worst localization (accurate only down to around 120 Hz, which does not cover any of the test signals). The listening test data confirms this prediction, as these locations are the worst performing in this case. The location predicted for best localization (down to around 50 Hz) is the left center location. Again, this agrees with the test data, as all test signals were accurately localized. The right center location, on the other hand, shows some disagreement. This location was predicted to have accurate localization down to around 80 Hz, however at 80 Hz in the test data, there is considerable localization error. This may be due to fire doors being located directly to the right of this location, which causes a slightly greater room width at this location. This issue is highlighted with much of the data for the right column of seats (but not replicated on the left side).

Source location 3's (source located near the corner of the listening grid, along the front of the test area) results are perhaps the least in agreement with the predictions (Fig. 4.5). The four furthest locations from the source seem to be roughly in line with predictions, but the left column of seats exhibit much greater accuracy than expected. A possible explanation for this is that, as previously noted, there was a considerable amount of large equipment stored at the front of the room (behind the source 3 location). This may have prevented the expected pattern of reflections predicted by the image source model, causing disagreement between objective and subjective data.

Looking at the data for source 4 (source located directly in front of the listening grid), it becomes apparent that the predictions based on the current theory may be slightly conservative for this particular configuration. The predicted minimum localizable frequencies (Fig. 4.2) for source 4 indicate that the front center listening location should give localization down to approximately 60 Hz. The listening test data, however, shows very accurate localization for all test signals, including the 40 Hz signals. This is also reflected for all centrally-located locations in the grid (Fig. 4.6). The rear corner locations, on the other hand, appear to be more closely in line with the predictions, giving the most precise localization only for the 115 Hz signals.

This variety of analysis could carry on for quite some time without drawing any definite conclusions, which would defeat the primary aim of this work. In order to inspect the listening test data in an efficient and reasonable manner, the data must be compressed into a single plot, if possible.

Since the intended outcome of this work is to determine what the minimum number of uncorrupted wavelengths needs to be received at a listening location to allow for accurate localization of a sound source, the listing data must be converted to be relative to uncorrupted wavelengths received vs. localization accuracy. Unfortunately, binaural room impulse response measurements weren't taken of the test room. In place of this, the image source model was used, keeping note that the actual test room isn't perfectly rectangular and contains some large pieces of equipment.

For each test frequency/source position/listener location combination the number of uncorrupted wavelengths was calculated. This was then analyzed based on how many correct localization judgments there were for each configuration. The allowance for correct localization was swept from $\pm 65^\circ$ down to $\pm 5^\circ$. If a listener's localization judgment was within the allowance range, then it was counted as a correct localization. After analyzing all the data, the percent correct localizations was calculated across the range of uncorrupted wavelengths received and plotted (Fig. 4.7).

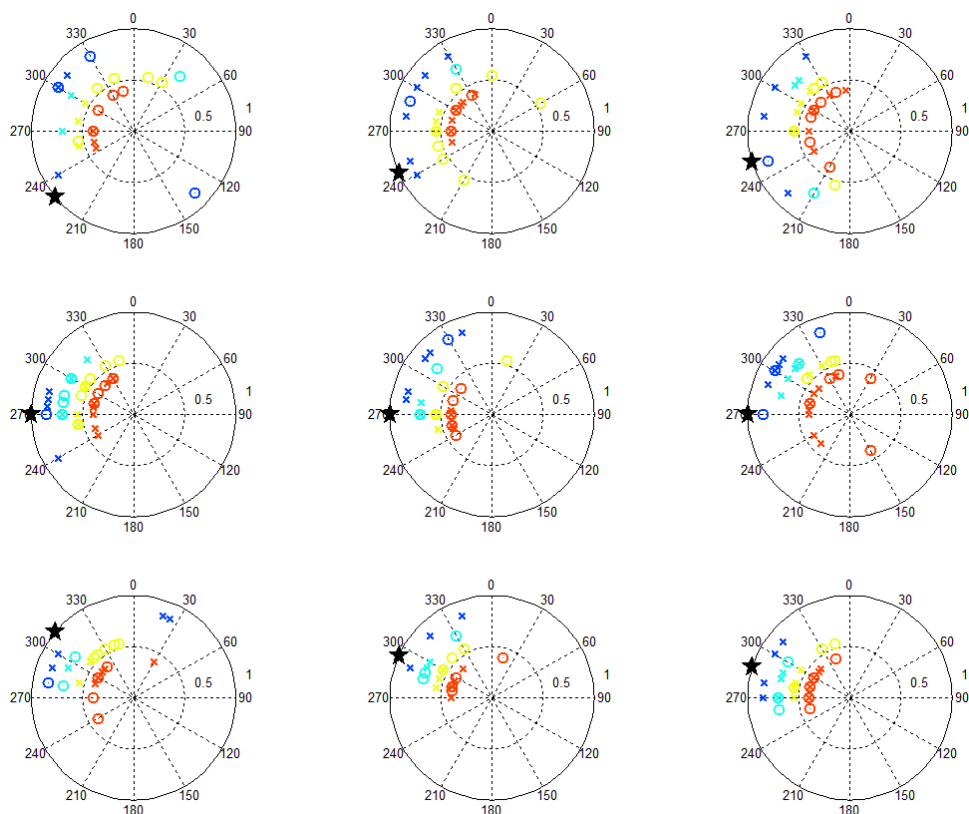


Fig. 4.3 Listening test results for all 9 seats (as laid out in Fig. 4.1) due to source location 1

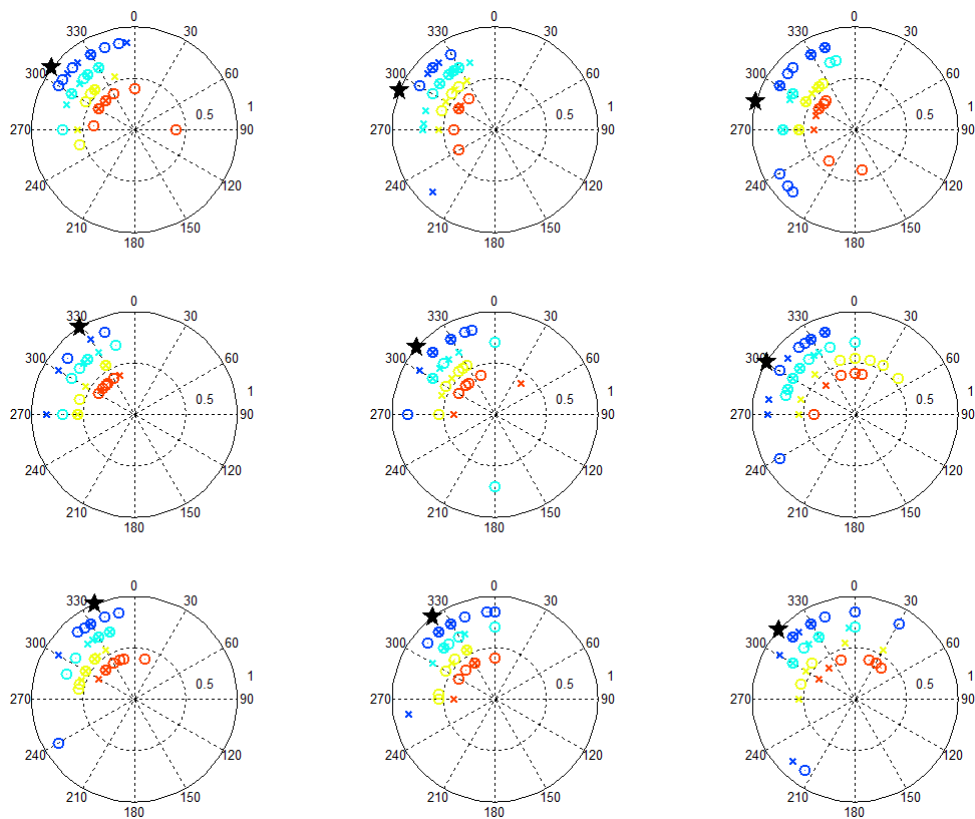


Fig. 4.4 Listening test results for all 9 seats (as laid out in Fig. 4.1) due to source location 2

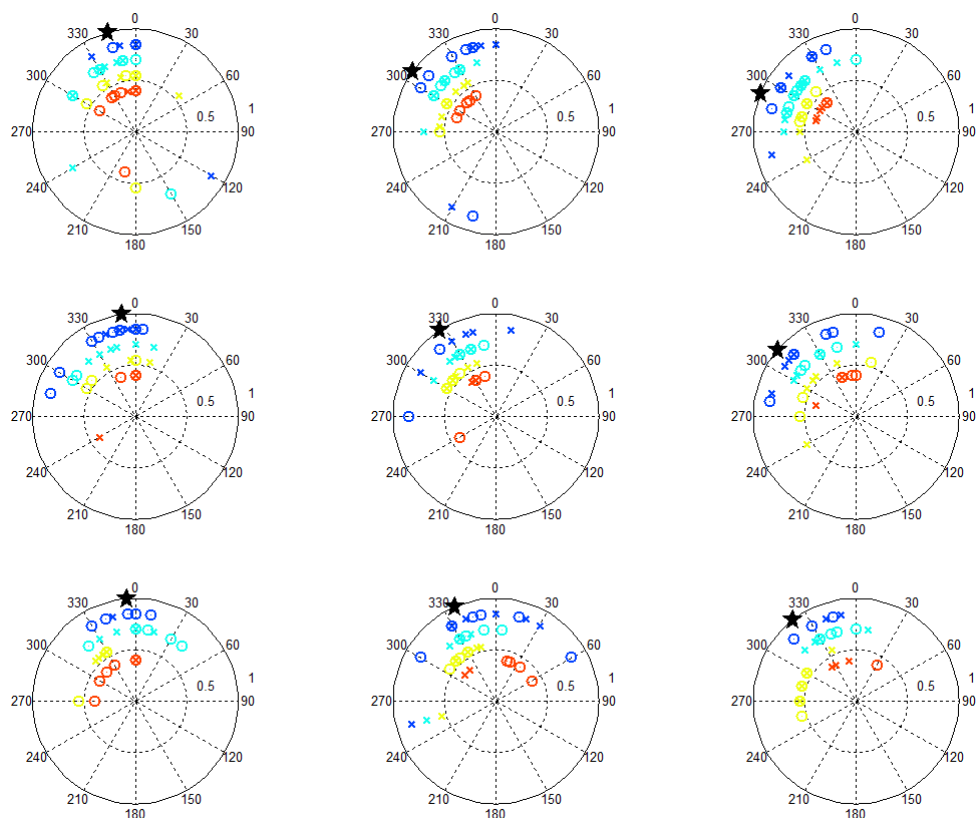


Fig. 4.5 Listening test results for all 9 seats (as laid out in Fig. 4.1) due to source location 3

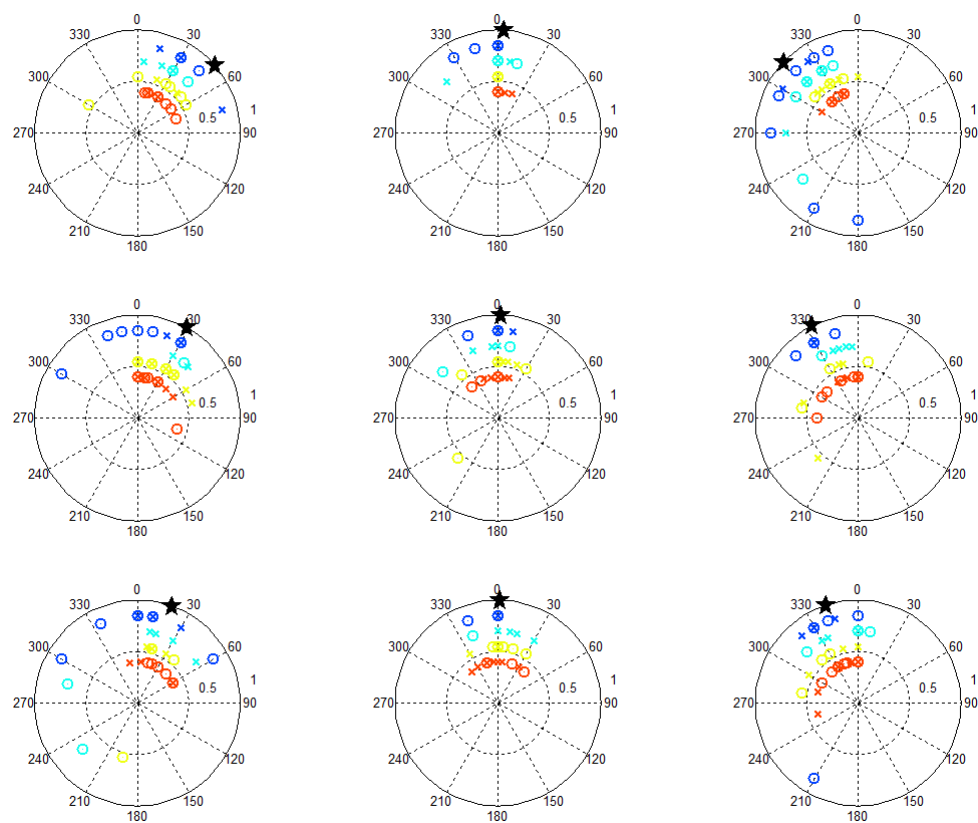


Fig. 4.6 Listening test results for all 9 seats (as laid out in Fig. 4.1) due to source location 4

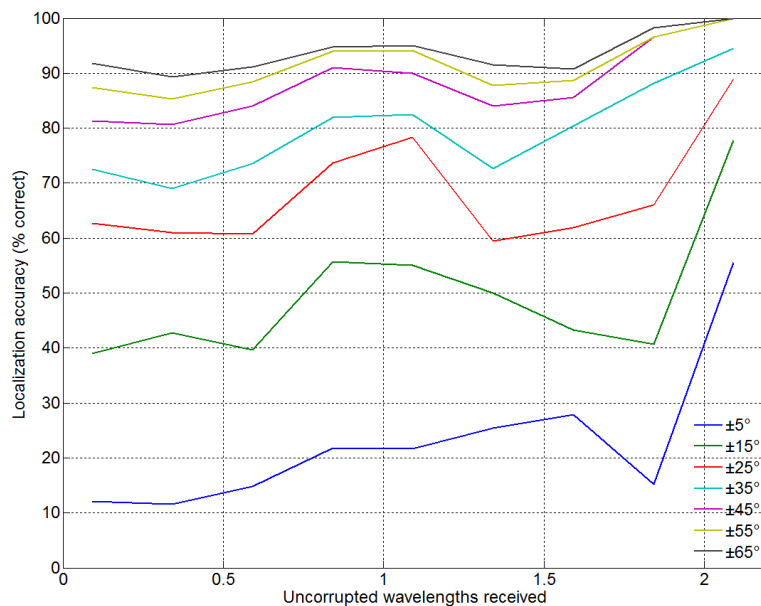


Fig. 4.7 Percentage of accurate localization judgments as compared to the predicted number of uncorrupted wavelengths received (traces represent different sizes of correct localization zones)

This form of analysis reveals trends in the data that could not be gleaned by manual inspection. Beginning with the broadest accuracy zone ($\pm 65^\circ$), it is clear that the number of uncorrupted wavelengths received plays little part in localization accuracy. This suggests that rough localization (to a general direction – left, right, front, rear) is possible regardless of the system configuration or source signal since the data shows over 90% localization accuracy for all cases.

As expected, the localization accuracy decreases as the zone contracts. At the smallest accuracy zone ($\pm 5^\circ$), the data shows around 10-20% correct localization until around 1.8 uncorrupted wavelengths have been received. At this point the accuracy sharply rises to around 55%. This trend can be seen for the broader accuracy zones, but is most pronounced for $\pm 5^\circ$.

This is a crucial finding in this research. The initial suggestion was that 1.4 uncorrupted wavelengths were needed to accurately localize a sound source. This was formulated by inspecting subjective evaluation results from previously published work (which did not test multiple seating locations and very few source locations). The subjective evaluation data from this thorough listening test indicates that the 1.4 wavelength recommendation needs to be adjusted to 1.8 wavelengths.

The updated theory of sound source localization in a closed acoustic space states, therefore, that in order to localize a given frequency accurately, a minimum of 1.8 uncorrupted wavelengths of the sound must be received.

4.4 Discussion of source signal characteristics

Up to this point, no distinction has been made between the windowed sine and square wave test signals. As the purpose of using both varieties of signal was to examine localization accuracy with and without additional harmonic content, it would appear necessary to keep analysis separate for these two signals.

Previous research indicates that sine waves or complex tones (harmonic series) both allow for the same localization accuracy¹³. It is thought that this is due to the signal's lack of strong trigger for the precedence effect. Noise signals, on the other hand, have been found to be much easier to localize, as the noise can be seen as a collection of sharp, impulsive signals, thus giving numerous triggers for the precedence effect. The data from the current listening test can be analyzed for the windowed sine and square waves separately to ensure this is the case (Fig. 4.8).

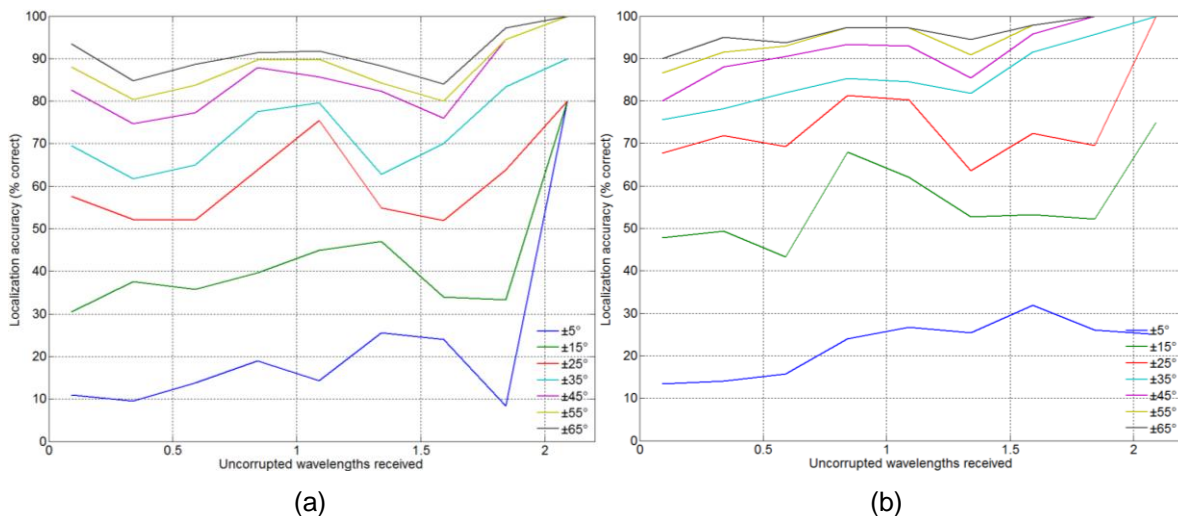


Fig. 4.8 Percentage of accurate localization judgments as compared to the predicted number of uncorrupted wavelengths received for windowed (a) sine wave and (b) pseudo-square waves

The trends in both sets of data are largely the same. Observable differences are that the square wave signals result in slightly higher localization accuracy and the sine wave with the narrowest correct localization zone ($\pm 5^\circ$) shows a sharp increase past 1.8 wavelengths received, while the square wave data shows no increase at all. This is likely down to noisy data, as there were not many data points for this particular scenario.

Overall, the data appears to confirm the notion that the sine and square waves can be considered together for the purposes of this work. Further investigation into this assumption will need to be carried out before this approach can be considered entirely appropriate for all cases.

5 CONCLUSIONS & FUTURE WORK

The research detailed in this paper sets out to confirm and to update an emerging theory of low-frequency sound source localization in closed acoustic spaces. Prior to this work, the theory was based solely on objective data analysis, thus requiring validation through subjective evaluations.

The resulting listening test data gives a strong indication that low-frequency localization is indeed possible in closed spaces. If only rough localization is necessary, then this should be possible regardless of system configuration. If precise localization is required, however, a listener must receive at least 1.8 uncorrupted wavelengths of the lowest frequency component in the signal.

This brings up the issue as to whether low-frequency directionality is necessary in the context of a broadband sound. As noted in this work, it is understood that incorrect low-frequency localization cues in the presence of accurate high-frequency cues can cause localization error. It is therefore recommended that if precise sound reproduction is desired that works universally whether or not high frequency content is present in the signal, directional information must be preserved across the whole frequency spectrum. This requires the elimination of the practice of employing a single subwoofer in surround sound systems or driving a subwoofer system at a large music festival with a monaural signal (the question of stereo subwoofer systems for large-scale sound reinforcement has been the focus of previous work by the authors¹⁷).

It must be stressed that the recommendation for 1.8 uncorrupted wavelengths for accurate localization requires further validation. The lack of BRIRs of the test room (and corresponding reliance on image source-generated approximate BRIRs) likely decreased the reliability of the analysis performed. A new headphone-based test is already underway at the time of writing, which is based on BRIR measurements in a small room and includes a variety of signals (continuous

sine/filtered noise and sine/noise bursts) which will overcome the limitations of this work due to the trouble of triggering the precedence effect with pure sinusoidal tones (with or without the corresponding harmonic series).

Despite the need for further validation, this work gives clear evidence supporting the notion that low-frequency localization is possible in closed spaces. This has been gathered from a thorough objective and subjective examination of the problem, considering all important variables (room topology, source/listener location, signal characteristics, absorption). Even in the context of broadband signals, accurate low-frequency directivity should avoid any localization confusion and will provide the best possible listening experiences for all.

6 ACKNOWLEDGEMENTS

The authors would like to thank Dave Imrie, Rob Bullen, Ed Culwick, Dave McGregor and Josh Small (final year students (all now graduates) of the BSc (Hons) Sound, Light and Live Event Technology program at the University of Derby) for their assistance in setting up, troubleshooting and conducting the listening tests.

7 REFERENCES

1. Hill, A.J.; M.O.J. Hawksford. "Low-frequency sound source localization as a function of closed acoustic spaces." Proc. Institute of Acoustics – Conference on Reproduced Sound, Manchester, UK. November, 2013.
2. Hill, A.J.; S.P. Lewis; M.O.J. Hawksford. "Towards a generalized theory of low-frequency sound source localization." Proc. Institute of Acoustics – Conference on Reproduced Sound, Brighton, UK. November, 2012.
3. Strutt, J.W. "On our perception of sound direction." Phil. Mag, vol. 13, pp. 214-232. 1907.
4. Jeffress, L. "A place theory of sound localization." Journal of Comparative and Physiological Psychology, vol. 41, pp. 35-39. 1948.
5. Zhang, P.X.; W.M. Hartmann. "Lateralization of sine tones – Interaural time versus phase." J. Aco. Soc. Am., vol. 120, no. 6, pp. 3471 – 3474. December, 2006.
6. Rakerd, B.; W.M. Hartmann. "Localization of sound in rooms V: Binaural coherence and human sensitivity to interaural time differences in noise." J. Aco. Soc. Am., vol. 128, no. 5, pp. 3052-3063. November, 2010.
7. Schnupp, J.; I. Nelken; A. King. *Auditory neuroscience: Making sense of sound* (Chapter 5: Neural basis of sound localization). MIT Press, Cambridge, MA, USA. 2011.
8. Rakerd, B.; W.M. Hartmann. "Localization of sound in rooms II: The effects of a single reflecting surface." J. Aco. Soc. Am., vol. 78, no. 2, pp. 524-533. August, 1985.
9. Haas, H. "The influence of a single echo on the audibility of speech." J. Audio Eng. Soc., vol. 20, no. 2, pp. 146-159. March, 1972.
10. Toole, F.E. *Sound Reproduction – Loudspeakers and Rooms* (Chapters 6 & 7). Focal Press, Boston, MA, USA. 2008.
11. Hartmann, W.M. "Localization of sound in rooms." J. Aco. Soc. Am., vol. 74, no. 5, pp. 1380-1391. November, 1983.
12. Rakerd, B.; W.M. Hartmann. "Localization of sound in rooms III: Onset and duration effects." J. Aco. Soc. Am., vol. 80, no. 6, pp. 1695-1706. December, 1986.
13. Rakerd, B.; W.M. Hartmann. "Localization of sound in a reverberant environment." Proc. 13th International Symposium on Hearing. 2003.
14. Hartmann, W.M. "Localization of sound in rooms: The effect of visual fixation." Proc. 11th ICA, Journal d'Acoustique, vol. 3, pp. 139-142. 1983.
15. Kuttruff, H. *Room Acoustics, 5th Edition*. Spon Press, London. 2009.
16. Hill, A.J.; M.O.J. Hawksford. "Visualization and analysis tools for low frequency propagation in a generalized 3D acoustic space." J. Audio Eng. Soc., vol. 59, no. 5, pp. 321-337. May, 2011.
17. Hill, A.J.; M.O.J. Hawksford. "On the perceptual advantage of stereo subwoofer systems in live sound reinforcement." 135th Convention of the Audio Engineering Society, paper 8970. New York, USA. October, 2013.