

LOW-FREQUENCY SOUND SOURCE LOCALIZATION AS A FUNCTION OF CLOSED ACOUSTIC SPACES

AJ Hill
MOJ Hawksford

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

1 INTRODUCTION

The general mechanisms used for the purpose of sound source localization have been well-researched ever since the work of Rayleigh in the 1800s¹. Over a century of research in this area has given a solid collection of localization rules that are generally accepted amongst researchers.

There remains confusion, however, regarding the topic of low-frequency sound source localization. Some experts hold the opinion that low-frequencies are unlocalizable in closed spaces²⁻⁷ while others claim localization is indeed possible, within reason⁸⁻¹³. Previous research by the authors¹⁴ begins the process of pulling back the veil on this topic, by developing a generalized theory on low-frequency localization.

This initial work towards a clear answer to the question of low-frequency localization puts forward a hypothesis that the lowest localizable frequency in a closed space is dependent on a number of variables including: room dimensions, source location, listener location and reverberation time. The work provides a good starting point for investigation into this area and poses a number of questions that remain to be answered¹⁴.

The research presented in this paper takes up the task of systematically answering these questions with the goal of strengthening the evidence of this generalized theory of low-frequency sound source localization.

The paper begins with a brief overview of existing knowledge (and debate) concerning sound localization while describing the findings from the initial work on this subject by the authors¹⁴. This is followed by an explanation of the simulation procedure and data analysis methodology used in this work, proceeded by the presentation of key simulation results and a comparison of these results to binaural measurements.

With the new findings concerning low-frequency localization capabilities with a single source, evidence is presented in relation to localization due to phantom imaging from two sources and how this lines up with the single source findings. The paper culminates in section 6 with conclusions from this research as well as suggestions for future work.

2 SOUND SOURCE LOCALIZATION

The foundation of our current understanding of sound localization was laid by John Strutt (a.k.a. Lord Rayleigh) in 1876¹. Rayleigh's duplex theorem describes the most important aural cues for localization; however the theorem fails to describe localization for all scenarios. Due to this, additional research has been carried out to resolve this ambiguity. This section briefly reviews sound localization and highlights the current understanding of low-frequency localization, emphasizing disagreement between previously published research findings.

2.1 General mechanisms

Current understanding of sound localization is rooted in Rayleigh's duplex theorem. The theorem describes two key mechanisms used in the localization of sound. First is interaural time delay (ITD) which operates on the fact that a sound arrives at each ear at a slightly different time (assuming the source isn't directly in front, behind, above or below a listener). This time difference (typically measured in the hundreds of microseconds) translates to a phase difference interpreted by the brain as the result of a source's location. ITD is an accurate cue up to around 1 kHz¹, although this is dependent on the angle of arrival (which is interesting as the cue's upper frequency accuracy limit is dependent on exactly what it's trying to measure!). Beyond this range, ITD becomes unreliable as multiple wavelengths fit between the two ears, causing confusion regarding phase difference.

The second key mechanism in Rayleigh's theorem is interaural level (or intensity) difference (ILD or IID). ILD picks up where ITD loses accuracy (around 1 kHz). While ITD operates on time of arrival differences between the two ears, ILD looks for received signal strength differences. In the high-frequency range, wavelengths are shorter than the dimensions of a typical human head, resulting in acoustic shadowing¹. This is illustrated by imagining a spotlight directly to the right of an individual. The person's right ear will be strongly illuminated while the left ear will be in the shadow of the head. The same effect occurs acoustically.

While the two mechanisms from duplex theorem go a long way to describe sound localization, some grey areas remain. Most notably, the theorem only operates on the front half of the horizontal plane, meaning that ITD and ILD struggle to distinguish between sounds arriving from front/back and above/below. Clearly, humans can distinguish these arrival directions, so there must be additional cues beyond ITD and ILD.

The ambiguity of ITD and ILD is largely resolved by another physically-based mechanism known as head-related transfer functions (HRTF). HRTFs operate on the fact that the human ear is asymmetrical; therefore, sound will take slightly different paths to reach the ear drum, depending on the angle of arrival. Very short reflections occur due to the structure of the ear, causing comb-filtering of the received signal. The characteristics of the comb-filtering are angle of arrival-dependent. This allows humans to determine fairly accurate direction of arrival¹⁵.

2.2 Present understanding of low-frequency localization

While the localization methods discussed in section 2.1 are largely accepted as correct, there are varied viewpoints concerning the localizability of low-frequencies (subwoofer band, typically below 120 Hz). A thorough literature review was conducted by the authors in a previous paper¹⁴ which indicated a nearly even split between those who believe low-frequency localization isn't possible or important²⁻⁷ and those who think the opposite⁸⁻¹³.

Of the papers reviewed, similarities emerged between all pieces of research. First, nearly all work used listening tests in a single room with a centrally-located listening point. Despite the similarities in experimental setups among the reviewed research, most offered different recommendations concerning low-frequency directivity. Some suggested that directionality isn't discernible below around 120 or 200 Hz^{2,6}, while others indicated that it was audible through the majority of the subwoofer band (nearly down to 20 Hz)^{8,9,11}. A number of the authors experimented with subwoofer location and found that placement directly to the sides of listeners showed the best localization performance^{4,6,10}.

Based on this review, it seems there has been little work focused on the effect of listener position and room dimensions on localization accuracy. Initial work focused on eliminating this ambiguity looked at the amount of uncorrupted localization time (time a listener receives direct sound only) at various locations in different two-dimensional spaces. Corner and front wall midpoint source positions were examined. The resulting rough hypothesis suggests that listeners require approximately 1.4 uncorrupted wavelengths of a given frequency for accurate localization and that position and room size were crucial factors determining the number of uncorrupted wavelengths received¹⁴.

3 LOW-FREQUENCY SOURCE LOCALIZATION ANALYSIS

In order to refine the initial theory developed in the previous work¹⁴, a more precise simulation procedure is needed to help develop robust equations concerning uncorrupted localization time as a function of room topology. Once the simulation points to a set of equations, the refined theory can be tested against binaural measurement data and examined alongside listening tests. This section will explore each of these key research areas and present results stemming from the simulations and experiments.

3.1 Simulation procedure

The authors' previous work in this area utilized a bespoke simulation toolbox which was built around a finite-difference time-domain (FDTD) modeling algorithm¹⁶. While shown to be quite accurate, the FDTD method requires long runtimes, which is why the previous work used only two-dimensional virtual spaces.

In order to move towards a verifiable theory on low-frequency localization, simulations need to be carried out for three-dimensional spaces. In addition, the previous work only examined 2 source locations (room corner and front wall midpoint) and nine listener locations. This restriction needs to be removed to allow for development of a robust theory.

With these requirements in mind, an image source acoustical modelling algorithm was developed from an existing piece of software previously used by one of the authors. The model allows for three-dimensional spaces and precise source and listener placement (where FDTD has limitations due to the finite grid point spacing).

Simulations were run in a virtual space of dimensions 5 m x 4 m x 3 m. A grid of 81 listening points was used where each point was spaced in intervals of 10% of the room dimension in question (with all points on the same horizontal plane). The source location was swept across the front of the room and then along one of the side walls, with each position spaced at a 20% interval of the room dimension under inspection. The remaining two walls can be assumed to give identical behavior to the two tested walls.

In addition to the source and listener position testing, room absorption was varied (5%, 10%, 20% and 60%) as well as listening height (30%, 60% and 80% of the overall room height).

The image source simulation was run for each source position, absorption level, listening height and listener location, where each listener point was run twice to gather data for the left and right ear, respectively (with 18 cm spacing). As the uncorrupted localization time was of interest, only the first 5 orders of virtual sources were included in the model, which reduced the simulation runtime.

3.2 Data analysis methodology

The simulated data was analyzed in an attempt to identify a trend of uncorrupted localization time due to the relevant variables, all in terms of typical localization cues explained by Rayleigh's duplex theorem.

The first stage of analysis windowed the generated impulse responses to eventually examine the phase response at various stages of the signal's propagation path. Hanning windows were generated beginning with a width of 75% of the theoretical propagation time to the listening point under examination and continuing in 0.2 ms steps until the propagation time for a signal to travel from one corner of the room and back had been reached.

With the windowed impulse responses in hand, the complex frequency response was calculated using a fast Fourier transform (FFT). The resulting phase response was used to extract the time delay at each frequency bin using Eq. 3.1. With time delays known for left and right virtual ears, the interaural time difference (ITD) could be calculated with Eq. 3.2.

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

$$ITD_{\text{received}}(f) = T_L(f) - T_R(f) \quad (3.2)$$

where $T(f)$ is the time delay (s) at frequency f , $d(\text{phase}(f))$ is the first derivative of phase at frequency, f , $d\omega$ is the spacing of frequency bins (rad/s), $ITD(f)$ is the received ITD at frequency, f , and $T_L(f)$ and $T_R(f)$ are the time delays (s) for the left and right virtual ears, respectively.

It is expected that as the time window applied to the impulse response is expanded, the error between the expected ITD (true source direction) and the received ITD (apparent source direction) will increase. The expected ITD (frequency-independent) is calculated using Eq. 3.3 (assuming height doesn't hugely affect the horizontal angle of arrival).

$$ITD_{\text{expected}} = \tan^{-1} \left(\frac{L_y - S_y}{L_x - S_x} \right) \quad (3.3)$$

where (S_x, S_y) and (L_x, L_y) are the source and listener horizontal locations (in meters), respectively. The ITD error can be converted to angle of arrival error using Eq. 3.4.

$$\theta_{\text{error}}(f) = \sin^{-1} \left(\frac{cITD_{\text{error}}}{r} \right) \quad (3.4)$$

where c is the speed of sound in air (m/s) and r is the approximate radius of a human head (taken as 9 cm in this case).

The angle of arrival error as a function of time was tracked for each listening location for one source location, absorption level and listening height at a time. From this data, the uncorrupted localization time (where angle of arrival error = 0 degrees) was determined and plotted as a function of listener location for each room configuration (example plot given in Fig. 3.1, with a corner subwoofer).

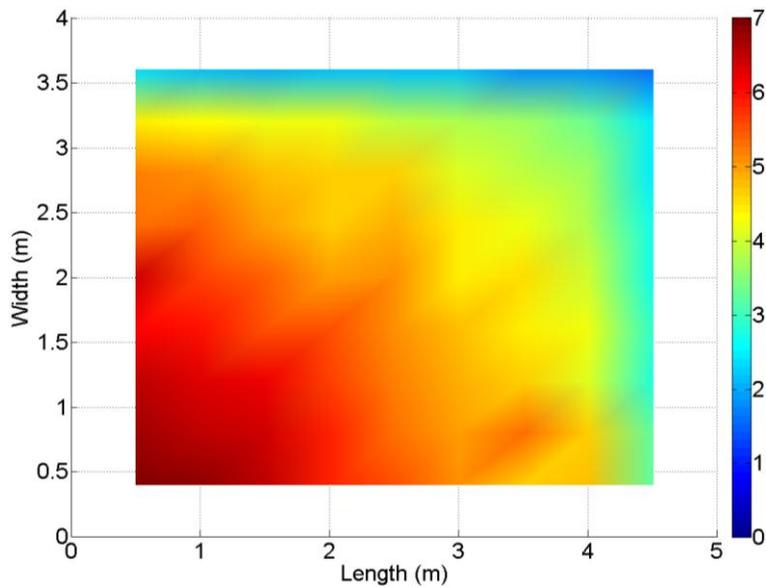


Fig. 3.1 Uncorrupted localization time (ms) due to a corner subwoofer on the floor of a 5 m x 4 m x 3 m room (5% absorption) with a listener height of 1.8 m

3.3 Simulation results

The assumption carried over from the previous paper on this subject¹⁴ is that the closer a listener is to a sound source, the longer that listener will have to properly localize the sound. This is apparently evident in Fig. 3.1, but upon inspection of non-symmetrical subwoofer placement localization time plots, this is shown to not be the case (Fig. 3.2).

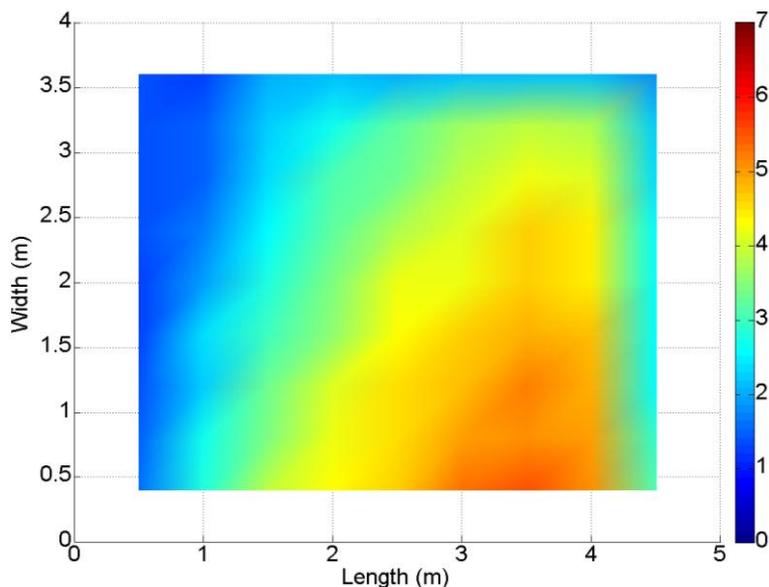


Fig. 3.2 Uncorrupted localization time (ms) due to a subwoofer at (1 m, 0 m) on the floor of a 5 m x 4 m x 3 m room (5% absorption) with a listener height of 1.8 m

In the case of a subwoofer positioned 1 m from the front wall of a room, the listeners closest to the source actually have a very short time to localize the source, while listeners further away along the length of the room (x-dimension) have the longest time. Upon careful consideration, it was determined that the localization time (as judged in this case by inspecting the phase response of the received binaural signals over time) was directly influenced by the time of arrival of the first reflection.

To confirm this idea, the original (unwindowed) impulse responses generated by the image source routine were analyzed by determining the time between the arrival of the direct sound and the first reflection. This time difference is the uncorrupted localization time and is plotted for the two scenarios previously discussed (Fig. 3.3). This data was generated using the Matlab script in Fig. 3.4.

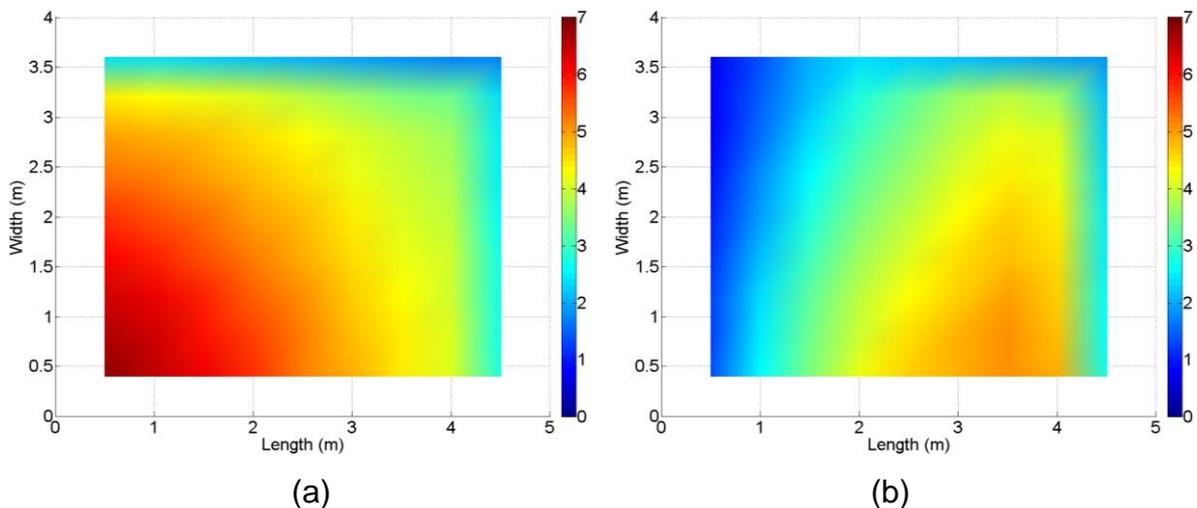


Fig. 3.3 Uncorrupted localization time (ms) (as determined by direct sound and first reflection arrival times) due to a subwoofer at (a) (0 m, 0 m) and (b) (1 m, 0 m) on the floor of a 5 m x 4 m x 3 m room (5% absorption) with a listener height of 1.8 m

It is clear that the time difference between the direct sound and first reflection arrival times is the crucial factor in defining the uncorrupted localization time in a closed space. All that is required to determine the localization time (which takes all geometrical configuration variables into account) is the arrival time of the first reflection.

```

1  % Lx, Ly, Lz = room dimensions (m)
2  % x, y, z = listener location indices (m)
3  % Sx, Sy, Sz = source location indices (m)
4  % c = speed of sound in air (m/s)
5
6  % first reflection arrival time storage variable (ms)
7  propT = 9999;
8
9  % calculate direct sound arrival time (ms)
10 Tdir = 1000*sqrt((x - Sx)^2 + (y - Sy)^2 + (z - Sz)^2)/c;
11
12 % loop to find propagation times for first-order reflections
13 for d = -1 : 2 : 1
14     if mod(d, 2) == 1
15         A = (d + 1)*Lx - Sx - x; % x-dimension
16     else
17         A = d*Lx + Sx - x; % x-dimension
18     end
19     for e = -1 : 2 : 1
20         if mod(e, 2) == 1
21             B = (e + 1)*Ly - Sy - y; % y-dimension
22         else
23             B = e*Ly + Sy - y; % y-dimension
24         end
25         for f = -1 : 2 : 1
26             if mod(f, 2) == 1
27                 C = (f + 1)*Lz - Sz - z; % z-dimension
28             else
29                 C = f*Lz + Sz - z; % z-dimension
30             end
31
32             % calculate the propagation time of the current reflection (ms)
33             T = 1000*sqrt(A^2 + B^2 + C^2)/c;
34
35             % see if the propagation time equals direct sound arrival ...
36             % time or is less than last found propagation time
37             if (T < propT) && (T ~= Tdir)
38                 propT = T;
39             end
40         end
41     end
42 end
43
44 % calculate the localization time (time difference between direct and first
45 % reflection arrivals)
46 locTime = propT - Tdir;

```

Fig. 3.4 Matlab code to determine the arrival time of the first reflection due to room dimensions and source/listener location

It is also essential to consider the average absorption in a room when determining the localization time. The time determined by the code in Fig. 3.4 assumes a room has perfectly reflective walls (zero absorption). As absorption is increased, the strength of reflections will decrease, eventually to the point where the space is anechoic, providing no reflections thus an infinite localization time.

Additional image source simulations were run with absorption levels ranging from 0% to 100% in steps of 10%. The resulting data was analyzed and a best-fit line equation was determined (Eq. 3.5 and Fig. 3.5, only partial data shown for clarity).

$$T_L(\alpha) = T_L(0) + \frac{e^{10\alpha}}{800} \tag{3.5}$$

where $T_L(\alpha)$ is the localization time for average absorption coefficient, α .

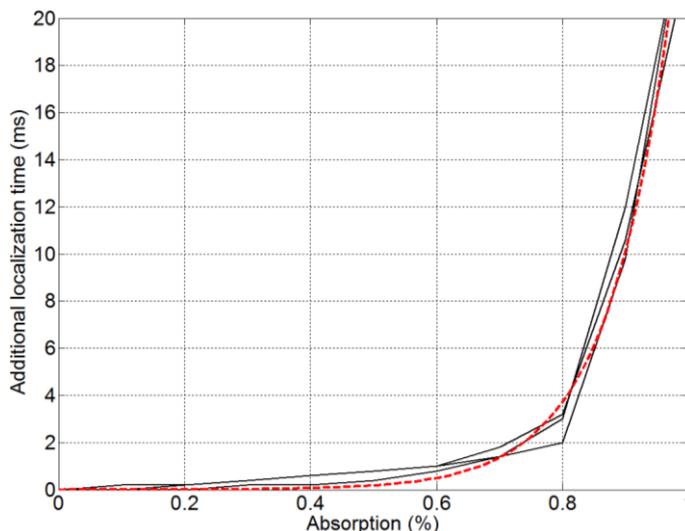


Fig. 3.5 Additional localization time versus absorption data (solid black lines) and best fit line (dashed red line) according to Eq. 3.5

Eq. 3.5 can therefore be used in conjunction with the code in Fig. 3.4 to determine the localization time for a given room configuration, assuming the space is rectangular. A modeling algorithm which lends itself well to non-rectangular topologies (such as FDTD) is needed to determine localization time for non-rectangular topologies. This is beyond the current scope of the research.

This absorption analysis leads to the idea that in order to increase absorption time in a closed space, the surface providing the first reflection (which is what corrupts directional cues) can be treated to increase its absorption, thus enhancing the localization ability at the given listening location(s). This is beyond the focus of this work, however, and will be left for future research.

3.4 Comparison to binaural measurements

In order to validate this technique, binaural measurements were taken in a listening room at the University of Derby. The room (dimensions 6.89 m x 6.30 m x 2.95 m) was arranged so that there was a 3 x 3 grid of measurement points at 1/4, 1/2 and 3/4 points along each horizontal dimension. Each measurement point had a height of 1.2 m. MLS measurements were taken using a dummy head and a subwoofer located along 1/4, 1/2 and 3/4 along both the front and left side walls of the room.

With the measurement data collected, the analysis routine described in section 3.2 was applied. The measurement data was then compared to the simulated data by plotting the measured localization times due to a corner subwoofer and a subwoofer $\frac{1}{4}$ of the way along the front room dimension (Figs. 3.6 & 3.7).

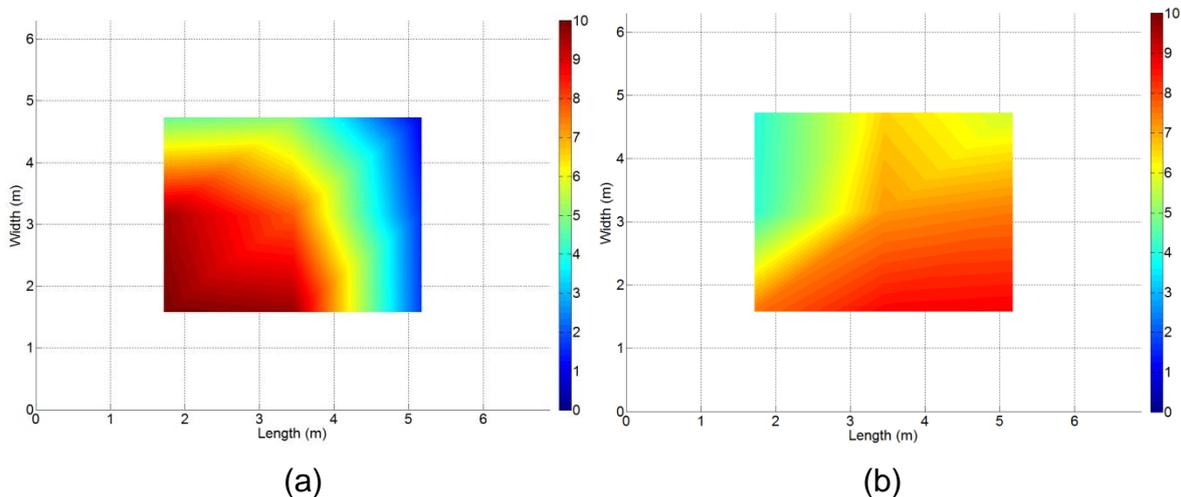


Fig 3.6 Measured uncorrupted localization times for a 6.89 m x 6.30 m x 2.95 m room with a subwoofer at (a) the room corner and (b) (0 m, 1.72 m)

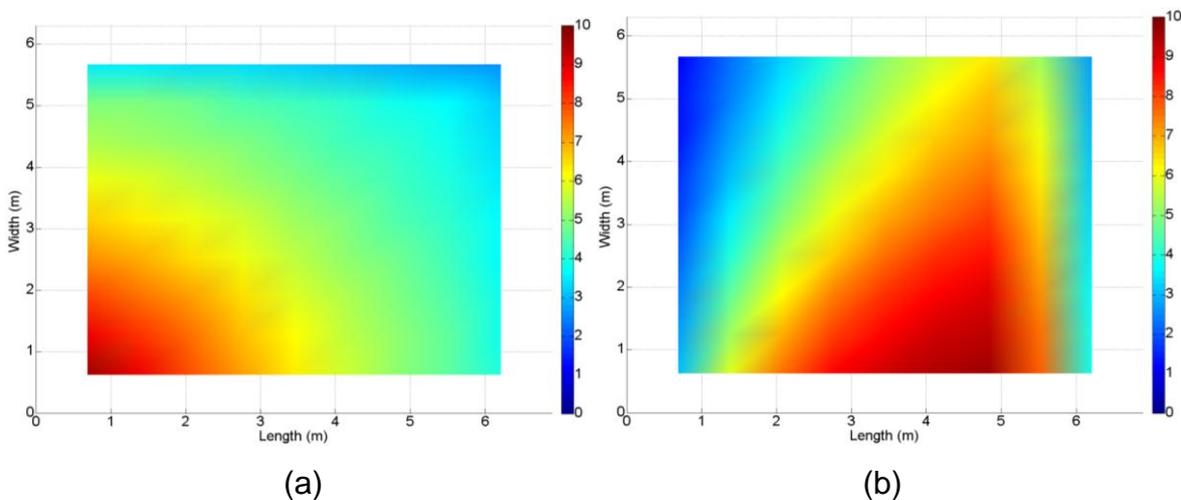


Fig 3.7 Simulated uncorrupted localization times for a 6.89 m x 6.30 m x 2.95 m room with a subwoofer at (a) the room corner and (b) (0 m, 1.72 m)

Comparing Figs. 3.6 and 3.7 indicate very similar trends between the simulated and measured data. Localization time peaks as a listener approaches the corner subwoofer, while localization time is maximized as a listener moves towards the rear third of the room length when the subwoofer is located 1.72 m from the front wall. These results indicate the accuracy of the proposed algorithm in determining the amount of uncorrupted localization time in a closed space.

It should be noted that there were some sources of error for the measurements. First, there was moderate noise in the corridor outside the listening room on the day of the experiments. Also, there was a considerable amount of furniture in the room, which may have caused early reflections to stray from the theoretical expectations. Even with these issues, the measurements are in line with the simulations, allowing the work to move forward with confidence of the methodology's accuracy.

4 APPLICABILITY TO PHANTOM IMAGING

An underlying aim of this track of research is to determine if monophonic subwoofer band reproduction in closed spaces is sufficient for adequate sound reproduction (as is currently assumed to be the case) or if multichannel low-frequency is needed to improve the listening experience.

Up to this point, this research has examined single source locations. As stereo sound reproduction relies on phantom imaging¹, it is essential to inspect the phase response, as in section 3.2, to determine if the same cues exist in stereophonic sound reproduction.

To prove the concept, impulse response data from the trials with a corner subwoofer at the front right and front left corners of the 5 m x 4 m x 3 m room were combined to give an impulse response reflective of the two sources operating in tandem. The resulting phase response was examined at each listening location and an apparent source location was determined. The uncorrupted localization times were plotted as in the previous examples (Fig. 4.1).

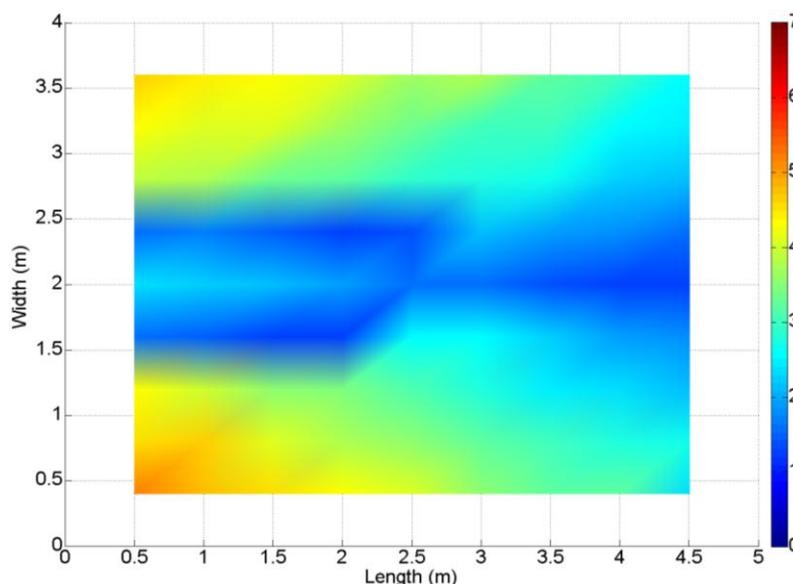


Fig 4.1 Uncorrupted localization time (ms) due to two corner subwoofers on the floor of a 5 m x 4 m x 3 m room (5% absorption) with a listener height of 1.8 m

The uncorrupted localization time for the stereo subwoofer system (Fig. 4.1) highlights some interesting points. Perhaps when first considering the phantom image's localization, it could be assumed that the localization time will exhibit a similar pattern as that of a physical source located at the phantom image location. Interestingly, this is not the case. The largest localization times are still in line with the individual subwoofer patterns, superimposed onto one another. This means that the reflections of the corner subwoofers, as expected, are still originating at the same locations and are therefore corrupting the localization cues at the same time after the direct sound's arrival.

Due to the two superpositioned patterns, the central "sweet spot" actually has the lowest localization time (although this spot will receive the most accurate phantom image for the high-frequency band). At low-frequencies, however, it seems that the best localization is off-center due to the greater difference between the direct sound and first reflection arrivals.

5 CONCLUSIONS & FUTURE WORK

Upon close inspection of sound propagation within closed spaces alongside binaural localization cues (as described in Rayleigh's duplex theorem), the issue of low-frequency localization has been clarified. This research highlights that the difference between the arrival of the direct sound and first reflection to a listener is the primary determinant of localization time. In addition, the average absorption of a space affects localization time, whereby high absorption coefficients (above 0.7) give multiple additional milliseconds of localization time.

Listening tests conducted in relation to this work have been informal in nature, but nonetheless reflect the objective findings of this paper and previous papers¹⁴. For this theory to be proven, however, detailed listening tests must be designed and carried out. Only with these tests will the idea of low-frequency localization as a function of room topology be conclusively proven.

Additionally, listening tests examining the effect of low-frequency directionality within the context of full-range sound reproduction must be conducted. Even if low-frequencies can be localized in certain spaces, it may not make any difference to the listening experience if the localization cues are dominated by the high-frequency band. Tests of this sort have been carried out in the context of live sound reinforcement¹⁷. The results from these tests indicate that directional low-frequency isn't perceptible in full-range systems, however, it aides in suppressing interference problems due to the decorrelation of the left and right signals.

Aside from the required listening tests, an objective examination into phantom imaging's effect on localization time is necessary as well as inspecting whether applying additional absorption to the surface causing the first reflection for a given listening location can help in increasing localization time.

Overall, the theory of low-frequency localization in closed spaces is now better understood. Where previous work makes gross generalizations (such as nothing under 100 or 200 Hz can be localized in any space), this research has shown that the minimum localizable frequency is a function of the configuration of a space (dimensions, source/listener location, absorption).

6 ACKNOWLEDGEMENTS

The authors would like to thank Haydon Cardew of the University of Derby for his assistance in taking the binaural measurements discussed in section 3.4.

7 REFERENCES

1. Rossing, T.D.; F.R. Moore; P.A. Wheeler. *The Science of Sound, 3rd Edition*. Addison Wesley, New York. 2002.
2. Borenus, J. "Perceptibility of direction and time delay errors in subwoofer reproduction." 79th Convention of the AES, paper 2290. October, 1985.
3. Welti, T.S. "Subjective comparison of single channel versus two channel subwoofer reproduction." 117th Convention of the AES, paper 6322. October, 2004.
4. Kelloniemi, A.; J. Ahonen; O. Paajanen; V. Pulkki. "Detection of subwoofer depending on crossover frequency and spatial angle between subwoofer and main speaker." 118th Convention of the AES, paper 6431. May, 2005.
5. Benjamin, E. "An experimental verification of localization in two-channel stereo." 121st Convention of the AES, paper 6968. October, 2006.
6. Zacharov, N.; S. Bech; D. Meares. "The use of subwoofers in the context of surround sound programme reproduction." 102nd Convention of the AES, paper 4411. March, 1997.
7. Kugler, C.; Theile, G. "Loudspeaker reproduction: Study on the subwoofer concept." 92nd Convention of the AES, paper 3335. March, 1992.
8. Griesinger, D. "Spaciousness and localization in listening rooms and their effects on the recording technique." JAES, vol. 34, no. 4, pp. 255-268. April, 1986.
9. Griesinger, D. "General overview of spatial impression, envelopment, localization, and externalization." 15th International Conference of the AES, paper 15-013. October, 1998.
10. Braash, J.; W.L. Martens; W. Woszczyk. "Modeling auditory localization of subwoofer signals in multi-channel loudspeaker arrays." 117th Convention of the AES, paper 6228. October, 2004.
11. Subkey, A.; D. Cabrera; S. Ferguson. "Localization and image size effects for low frequency sound." 118th Convention of the AES, paper 6325. May, 2005.
12. Martens, W.L. "Subjective evaluation of auditory spatial imagery associated with decorrelated subwoofer signals." Proc. 2002. International Conference of Auditory Display. July, 2002.
13. Beaubien, W.H.; H.B. Moore. "Perception of the stereophonic effect as a function of frequency." 11th Convention of the AES, paper 123. October, 1959.
14. Hill, A.J.; M.O.J. Hawksford. "Towards a generalized theory of low-frequency sound source localization." Proc. IOA – Reproduced Sound 2012, Brighton, UK. November, 2012.
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." *Journal of the Audio Engineering Society*, vo. 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 AES, New York. October, 2013.