How do the Characteristics of a Sound Event, Impact our Proficiency and Timing of Accurately Localising it?

Oliver McIntyre

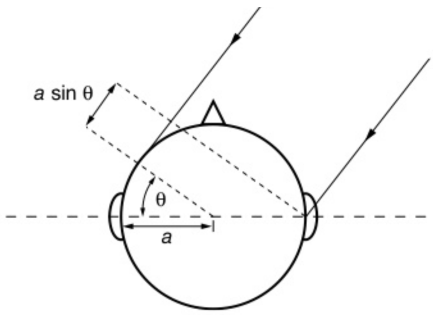
In this work I aim to study how interaural times cues, that help us localise sound, vary depending on the type of signal present.

Phenomena Relevant to Study

'Spatial Localization' as a term itself is defined by Blauert as "the location of an auditory event, which is related to a specific attribute or attributes of a sound event". Put more simply, one could say that spatial localization is a listener's capability of detecting the direction and distance of a sound source. In this study however, the term sound lateralisation rather than localisation would more aptly describe the circumstances of the experiment, as it will be undertaken using headphones. According to Moore (2004, p233), "headphones allow precise control of interaural differences and eliminate effects related to room echoes. Thus, lateralisation may be regarded as a laboratory version of localisation'.

The principal theory on localization was initially put forward by British physicist Lord Rayleigh, who proposed the 'Duplex theory' in his 1894 publication *'The Theory of Sound'*. Rayleigh's theory suggested that the auditory system localized sound by using binaural cues. These 'cues' Rayleigh proposed were Interaural Time Differences (ITD), and Interaural Level Differences (ILD). Simply put, they were based on the perceived difference in signals between the left and right ears. ITD are as a result of the fact that there are path differences from a sound source to each ear, and as a result, there is a disparity between the arrival times of that same sound to each ear. Blauert stated "dissimilarities between the two ear input signals related to the time when the signals occur". ILD are caused by the shadow of the head. For example, if a sound source is located to the right, the sound at the left ear will be lower in level than the right. In this study however, we will be focussed purely on ITD.

Figure 1



Looking at figure 1, it is shown that the path distance from a sound source in the top right (of the azimuth plane) to the left ear is greater than the distance to the right, and so the signal will arrive at the left later then the right. The difference in these times is the ITD. Therefore, in order to explore our efficiency of locating sound sources using ITD over headphones, I need to artificially emulate this difference in arrival times. This is done by playing two identical signals through each channel of the headphones, and gradually applying a delay to one of those channels. As the delay of one channel increases, so does the perceived location of the sound.

The precedence effect is a collective term to describe the effect of when two sound signals arrive within a sufficiently short time frame, so that they are perceived as a single sound event, but who's location is dominated by the first signal's location. Originally coined by Wallach in 1949, it is said to become prominent if the delay between the two signals is somewhere in the range of 1ms and 50ms. In a study into sound location precision by Ruth Litovsky, she remarks that studies into the precedence effect under headphone conditions result in "Lateralization of the compound (lead lag) stimulus depends primarily but not solely on information from the first arriving wavefront." Therefore as the delay time increases, the participants should be increasingly able to localise the source.

If the delay between the two signals is less than 1ms however, then an effect called summing localisation takes over. This is essentially an averaging between signals. For example, if a subject sat in front of two loudspeakers, each playing an identical sound signal with a delay between them of anything less than 1ms, then the signals are 'added' or 'summed' together, and the sound is perceived as coming from a location centrally between the two speakers. Additionally to this, William Hartman notes in *Binaural and Spatial Hearing in Real and Virtual*

How do the Characteristics of a Sound Event, Impact our Proficiency and Timing of Accurately Localising it?

Oliver McIntyre

Environments (p192)*,* that 'within this regime (1ms) there is an orderly weighting: the greater the delay, the less the weight given to the second click in the averaging'.

If the intensity or delay of one of two identical sources in altered past the point of summing localisation, then the phenomenon that is fusion will assume control. Handel describes fusion in depth:

> "Imagine two audio speakers separated by a medium distance. If the sound through each speaker is identical, then all the sound will appear to come from a point between them (summing localisation). If we adjust the intensity and timing slightly, the sound moves between the speakers toward the earlier, more intense one. If the intensity and timing differences are increased further, then all the sound appears to come from one speaker".

By using a wide range of delay times in my experiment, from 0ms to 0.66ms, I hope to explore to what extent the aforementioned phenomena operate and how they vary depending on the type of signal. I also hope to determine how the maximum delay before a source is localised varies depending on the signal type.

Lastly, it is important to mention localisation blur. This is the accuracy of identifying changes in the location of sound sources is measured as the minimum audible angle (MAA) or "localization blur". Localisation blur is, principally "the precision by which the location of a sound image can be given" (*Psychoacoustics: Facts and Models)*.

Investigation Aim

To determine how the characteristics of a sound event, specifically its frequency content, and whether it is continuous or pulsating, impacts our proficiency of accurately localising it. By gradually increasing the delay of one channel of two identical sound signals, I can emulate the ITD that would be experienced were the sound source legitimately located on the azimuth. By recording the delay time between the two channels that is required for the subjects to localise each signal type, I can determine which type of signals that are most easily localised.

Participants

8 Younger adults (6 men, 2 women, mean age: 20 years, age range: 19-23 years), all University of Kent students. They were recruited from Bridge Warden's College on the Chatham Dockside campus. Prior to the experiment, all participants were confirmed to have good hearing with no perceptible defects. They were all considered naïve participants.

Procedure

One single pair of headphones were used for the entire experiment for overall consistency and to ensure an identical signal level and frequency response was experienced by each participant. The headphones used were a pair of Sennheiser HD 201 closed dynamic stereo headphones.

The experiment was run by a MAX 7 patch on an iMac computer. The patch controlled the start and stop function of the signal, the type of signal, and the gradual delay increase of the signal. Before starting the experiment, it was explained to each participant on how the patch was controlled, so that it could be operated by them individually. It was also explained that the patch must be stopped by the participant themselves immediately after detecting the locality of the sound source, so that the delay required to localize were was as accurate as possible.

Each participant was subjected to eight different sound signals, which varied in length and frequency content. They were as follows; sinusoidal continuous, sinusoidal pulse, saw wave continuous, saw wave pulse, white noise continuous, white noise pulse, pink noise continuous, and pink noise pulse. For each of the pulsating signals, an impulse speed of 300ms and impulse length of 200ms was decided upon. For both the sinusoidal and saw signals, a frequency of 1000 Hz was also agreed. To ensure consistency, a volume level within the max patch was kept constant throughout, to ensure that volume level could not impact the ability to localise. The volume level of the mac was also kept constant.

Prior to the experiment, each of the eight types of signals were loaded into the max patch in the same order as shown on the experiment hand-out. The participant would start the patch, and instantaneously the specific type of signal would begin. The delay of one channel would immediately begin to be delayed from 0ms to 0.6ms over the course of 10 seconds. As soon as the subject localised the sound source they would hit the spacebar, which would instantly stop the patch. I would then record the delay time that was required, and ask the participant whether they perceived the sound as coming from the right or the left. I would then take note of the actual direction and compare it with theirs. Figure 2 shows the data sheet used to collect data from the participants.

Name of Participant…………………………………………………… Gender……………. Age……… Experienced/Naïve Participant?...

- You will hear 8 separate signals played one at a time.
- Each sound is initiated by the participant themselves by pressing the 'space' key
- Each sound will begin in the center and gradually move either left or right
- Once you are confident you have determined which way the sound is moving, immediately press the 'space' key once again.
- The assessor will then ask the participant which way they perceived the sound to move.
- The assessor will then take note of the direction the participant localised the sound, the actual direction the sound was coming from, and the delay between the two channels that was required for the participant to localise that sound.
- The participant is then free to start the process again for each of the remaining signal types.

Type of Signal	Direction of Localisation	Actual Direction (filled in by assessor)	Delay Time Required (ms)
Sine Wave – Continuous			
Sine Wave – Pulse			
Saw Wave – Continuous			
Saw Wave – Pulse			
White Noise – Continuous			
White Noise – Pulse			
Pink Noise – Continuous			
Pink Noise – Pulse			

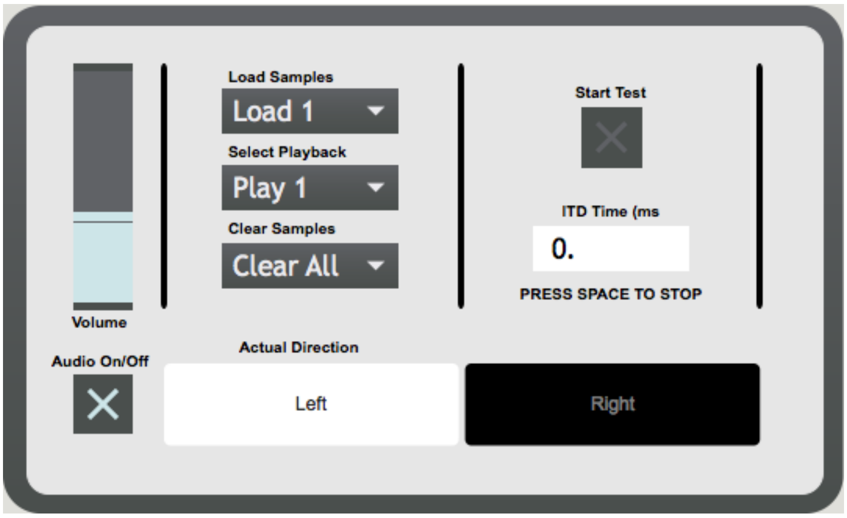


Figure 3

Figure 3 shows the patch used to run the experiment. In the bottom left is the audio on/off button, this was switched on by the assessor. On the left hand side is the volume slider. Before loading the samples into the patch I adjusted the levels of each of them to ensure the perceived loudness was as consistent as possible. This was to ensure that certain signals would not be localised better purely because they were perceived as louder. The volume slider therefore was kept constant throughout. In the centre of the patch are the 'load', 'play' and 'clear' functions. All 8 signals could be loaded into the patch beforehand to ensure a smooth running of the experiment without delays. Once loaded in, the participant is free to press the spacebar to initiate the patch and again to stop it once the sound has been localised. The assessor would then select ‘Play 2’ to move onto the next signal type and the process could begin again for the remaining signals.

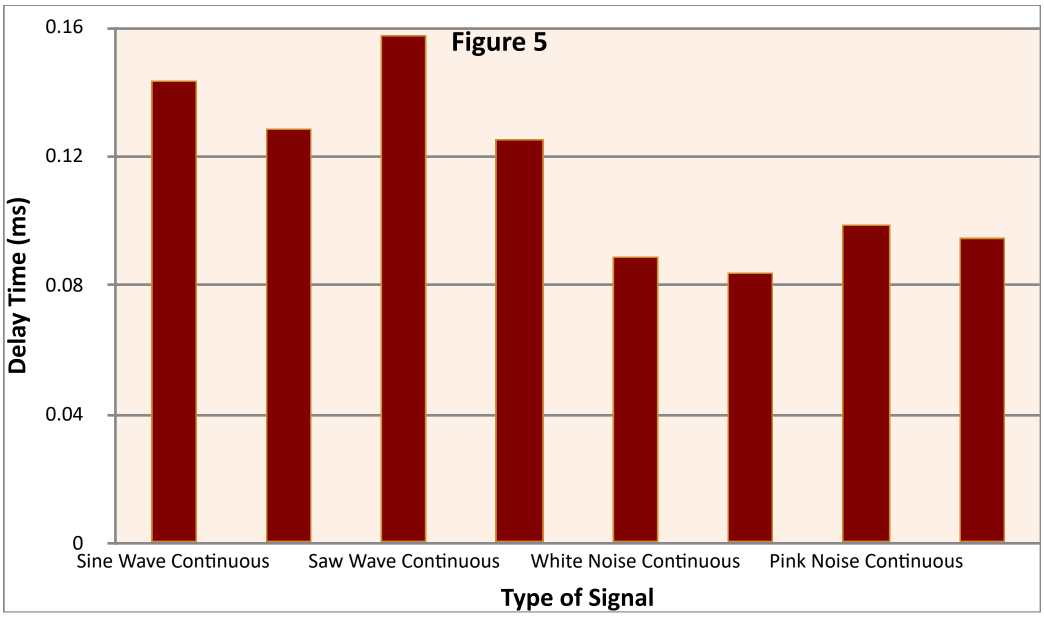
Results

Figure 4

Type of Signal	Average Delay Required (ms)	Median Values (ms)
Sine Wave – Continuous	0.1437	0.1439
Sine Wave – Pulse	0.1290	0.1274
Saw Wave – Continuous	0.1572	0.1551
Saw Wave –Pulse	0.1256	0.1234
White Noise – Continuous	0.0886	0.0891
White Noise – Pulse	0.0839	0.0951
Pink Noise – Continuous	0.0988	0.0931
Pink Noise - Pulse	0.0947	0.0924

Analysis of Results

From my results, it is shown overall, that noise signals have a lower average delay time required to localise than that of the sine and saw signals. This suggests that noise signals are most easily localised than that of sine and saw signals. The results between white and pink noise are pretty similar, with white noise marginally quicker. Saw waves have the longest delay time showing that they have the highest localisation blur when compared to noise signals. Also please note, that for each type of signal, the pulsating version of that signal has a shorter delay time across the board. This shows that when a signal is presented to a subject as a pulse rather than a



How do the Characteristics of a Sound Event, Impact our Proficiency and Timing of Accurately Localising it?

Oliver McIntyre

continuous sound, the delay time required is less, and it is more easily localised. Looking in more detail, it is shown that when sine and saw signals are presented as a pulse, the delay time required to localise is reduced significantly. For saw waves, the delay time for a pulsating version of the signal is roughly 20% shorter than that of the continuous, and for sine waves 10% shorter. Pulsating versions of the noise signals also reduce the delay time, but not as notably.

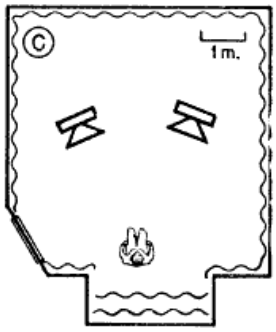
Comparisons of my Results with other Publicised Studies and Collected Information

Blauert describes (page 97), our proficiency of localising noise signals in comparison with other types of signals:

'It was evident that the discrepancy between the direction of the sound source and that of the auditory event was not random; there were constant tendencies for specific errors to occur with specific types of signals. With broadband signals, especially those of long duration or those repeated several times, a quite good agreement between the direction of the sound source and that of the auditory event was usually observed.'

As mentioned in the analysis, my results showed that proficiency of localisation in white and pink noise signals was superior to that of sinusoidal and saw waves. This suggests that signals with a greater frequency spread are more easily localised. This was true for both pulsating and continuous signals. Therefore my results reinforce Blauert's findings.

Figure 6



A study into The Precedence effect in sound localisation by Hans Wallach (1949, p318) was conducted that showed results that strongly support my own. Wallach's experiment demonstrated the localisation effect by using two loudspeakers rather than headphones. Shown in Figure 6, Wallach arranged two loudspeakers at equidistant from the subject, and played two identical signals driven by the same voltage, but with one delayed from 7ms to 70ms. All participants localised the sound as coming from the speaker that was not delayed. In a similar experiment, this time with a phonograph record, Wallach documented that when the sound was delayed in one speaker by 35ms, the sound was "localised by all subjects in the direction of the speaker which was ahead in time". (p318) As all participants in my study correctly identified the direction of the sound signal when one channel was delayed, it shows that ITD is an effective cue for localising sound.

In a study by Dustin Lindley, he intends to find out to what degree signals with a broad range of frequencies are localised superiorly to a pure tone. In his experiment he uses a voice, which is rich in transients. He states that "In nature, most every sound that could signify some sort of danger to us or prey for us would not be a pure tone, but would be very rich in transients. It makes sense that we can localize these sorts of sounds more effectively than pure tones, which really do not exist outside of human creation.' He finds that for the signal with many transients, "the brain can discern a difference in arrival time of 0.045ms". He also states "For pure tone signals (sine waves), the brain has a hard time localizing at all, regardless of the delay, unless the signal is moving across the auditory field of view." My findings agree with these results as the delay time required to localise for both sets of noise signals (with broad frequency range), was significantly shorter than that of the pure sine and saw signals.

Conclusion

In conclusion, I believe my experiment was successful in agreeing with theories and studies concerning signal types and localisation cues. For future experiments, several improvements could be made. Firstly, using a greater number of participants would dramatically increase the credibility of the data, particularly when the difference in delay times is so minute. Secondly, the nature of my experiment always gives the participant a 50/50 chance of correctly identifying the direction of the signal – either left or right, with the possibility that they do not actually perceive it. So in future studies I would use more precise directional questions, such as degrees around the head to eliminate this problem. I would also use a broader range of signal types in future studies. It is well documented that pure tones of high frequencies (above 1.5kHz) are much more difficult to localise that those around 500Hz, so in future tests, I would use sine and saw wave signals at a broad range of frequencies in order to explore this area more acutely.

Bibliography – Books

BLAUERT, J. 1996 *'Spatial Hearing: The Psychophysics of Human Sound Localization – Revised Edition'*. London: The MIT Press.

GILKEY, R. H. and ANDERSON, T. R. 1997. *'Binaural and Spatial Hearing in Real and Virtual Environments'*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

MOORE, B. C. J. 2004. *'An Introduction to the Psychology of Hearing, Fifth Edition'*. Bingley, Bradford: Emerald Group Publishing.

RAYLEIGH, J. W. S. 1894. *'The Theory of Sound'*. London: Macmillan and Co. Ltd.

RUMSEY, F. 2001. *'Spatial Audio'*. Oxford: Focal Press.

SCHNUPP, J. and NELKEN, I. and KING, A. 2011. 'Auditory Neuroscience, Making Sense of Sound.' Cambridge, Massachusetts: The MIT Press

HANDEL, S. 1993. *'Listening, an Introduction to the Perception of Auditory Events'*. Cambridge, Massachusetts: The MIT Press.

Bibliography - Websites

DIZON, R. M. and LITOVSKY, R. Y. 2004. *'Localization Dominance in the Media-Sagittal Plane: Effect of Stimulus Duration'*. The Binaural Hearing & Speech Lab.

http://www.waisman.wisc.edu/bhl/about_publications/2004DizonLitovskyJAcoustSocAm.pdf

LITOVSKY, R. Y. and COLBURN, H. S. and YOST, W. A. and GUZMAN, S. J. 1999. *'The Precedence Effect'*. Department of Electrical Engineering – Columbia University.

http://www.ee.columbia.edu/~dpwe/papers/LitovCYG99-precedence.pdf

LITOVSKY, R. Y. and MACMILLAN, N. A. 1994. *'Sound Localization Precision under Conditions of the Precedence Effect: Effects of Azimuth and Standard Stimuli'*. The Binaural Hearing & Speech Lab.

http://www.waisman.wisc.edu/bhl/about_publications/1994LitovskyMacmillanJAcoustSocAm.pdf

BROWN, A. D. and STECKER, G. C. and TOLLIN, D. J. 2014. *'The Precedence Effect in Sound Localisation'*. Journal of the Association for Research in Otolaryngology. Springer Link.

http://link.springer.com/article/10.1007%2Fs10162-014-0496-2#page-1

BROWN, A. D. 2012. *'Temporal Weighting of Binaural Cues for Sound Localization'*. The University of Washington.

How do the Characteristics of a Sound Event, Impact our Proficiency and Timing of Accurately Localising it?

Oliver McIntyre

https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/21842/Brown_washington_0250E_10994.pdf?sequence=1&isAllowed=y

WALLACH, H. and NEWMAN, E. B. and ROSENZWEIG, M, R. 1949. *'The Precedence Effect in Sound Localization'* from The American Journal of Psychology – Vol. 62, No. 3. JSTOR.

http://www.jstor.org/stable/1418275?seq=1#page_scan_tab_contents

FASTL, H. and ZWICKER, E. 2006. *'Psychoacoustics, Facts and Models: Third Edition'*.

http://www.autistici.org/2000-maniax/texts/Psychoacoustics%20-%20Facts%20And%20Models%20-%20Third%20Edition%20%28Hugo%20Fastl%29.pdf

LINDLEY, D. 2004. *'Experiments in Sound Localization'*. Department of Physics, Illinois University.

https://courses.physics.illinois.edu/phys406/Student_Projects/Fall00/DLindley/DLindley_Final.pdf

LITOVSKY, R. Y. and RAKERD, B. and YIN, T. C. T. and HARTMANN, W. M. 1997 *'Psychophysical and Physiological Evidence for a Precedence Effect in the Median Sagittal Plane'.* Department of Physics and Astronomy, Michigan State University.

https://www.pa.msu.edu/acoustics/litovsky.pdf