Bell vs yell: comparative ease of localisation of acoustic warning strategies for bicycles

BSc Sound Engineering and Production final year dissertation project

Allison Edwards (00719304)

Supervisor: Trevor Cox
Second reader: Rebecca Vos

May 2026

1 Abstract

This study investigated the ease of localisation of two acoustic warning strategies for bicycles, namely voices and bells. Recordings of voices in two vocal registers were tested, as well as two types of bell. Two listening tests were performed. The first was a small scale pilot test, primarily intended as a shakedown for the methods planned for the full scale test. The second test was a full scale localisation test conducted with a circular array of 8 loudspeakers and Ambisonics background noise. In the full scale test, time-to-identify (the time taken from onset of a sound to a listener making a localisation judgement) was also measured. It was found that voices provide much better localisation accuracy than bells, in agreement with conventional theory. However, it was also found that one of the bells under test (a multiple-impulse bell with a rotary internal mechanism that strikes the body of the bell multiple times) showed significantly lower time-to-identify, despite its lower localisation accuracy. Several avenues for further study are also proposed at the end of the report, such as testing specifically reaction time rather than response latency, and giving listeners tasks to perform during tests.

2 Acknowledgements

I would like to thank:

- Ren for providing the “female register” voice
- Henric and Rowan for putting up with my nonsense

3 Table of contents

3.1 Table of figures

4 Introduction

Bicycles are a common sight on city streets. In 2024, 41% of the population of England owned a bike, and 13% of the population of England cycled at least once a week (cyclinguk.org). Bikes often share space with pedestrians, particularly in pedestrianised areas of city centres, and so avoiding collisions between bikes and pedestrians is important. The commonly used strategy is an audible warning, usually a bell. The UK's Highway Code states "it is recommended that a bell is fitted to your bike" (Department for Transport, 2025, Rule 63), and UK law requires new assembled bicycles to be sold with a bell (The Pedal Bicycles (Safety) Regulations 2010), but does not require a bell to be on the bike after that point.

A point of note, and in fact the foundational reason for this study, is that when trying to avoid collisions between cyclists and pedestrians, it is not enough for pedestrians to simply know that there *is* a bike, they must also know *where* the bike is - and plenty of anecdotal evidence suggests that bicycle bells are quite difficult to localise. The obvious alternative is speech. Every bike has a rider, and the vast majority of riders will be capable of speech. In fact, the Highway Code also recommends politely informing pedestrians of one's presence

with your voice, as a replacement or augment to a bell. By taking a quick look at the respective spectrums of bells and speech, one finds that bells are extremely narrow-band, whereas speech contains energy over a much wider portion of the audible spectrum. All conventional wisdom and accepted theory would suggest that the anecdotal evidence is correct, as broad-spectrum sounds provide far more information for the human localisation system to work with and as such are much easier to localise. A point of comparison is the trend in recent years for large vehicles to use white or pink noise to notify nearby people that they are moving or reversing, rather than the more traditional “backup beeper” (Vaillancourt, 2013).

Another interesting factor to analyse is time-to-identify. This is related to, but distinct from, reaction time. In the present study, reaction time was considered to be the time it takes a listener to register that a bike is present, and was not measured or analysed.
Time-to-identify, however, is defined here as the time it takes a listener to come to a decision (correct or not) on which direction a bike is approaching from. This is different from reaction time, as a listener can be cognisant of the presence of a bike well before they have made a localisation judgement.

5 Theory

5.1 Literature review

Jin et al (2002) discussed the localisation of speech. They performed a localisation test using a binaural reproduction setup, having measured all participants’ HRTFs beforehand with ear canal microphones in an anechoic chamber. They found that broadband noise is very easy to localise, and broadband speech is slightly harder because the spectrum varies with time. However, speech does contain plenty of low and high frequency components which aid localisation. In a listening test, they found that cone of confusion errors (eg. front-back reversal) occurred more often with speech than with noise.

Withington (2019) found that emergency vehicle sirens have poor localisability as they only contain frequencies within a narrow bandwidth, roughly 500Hz to 1.8kHz. She tested several existing siren designs and a range of novel siren designs in a driving simulator. In the test, 8 loudspeakers were positioned around the simulator car, and a response panel with 16 positions was mounted in the car. The additional positions allowed greater resolution when determining the localisation errors made. The test found that the existing sirens had very poor localisation performance, whereas the new designs, which contained rising frequency sweeps interspersed with blasts of broadband noise, were localised correctly much more often. Subsequent road tests confirmed the validity of the lab tests. Emergency vehicles using the new siren designs found that drivers identified the direction of the oncoming vehicle faster and as such could give passage more easily, leading to the emergency vehicle moving faster on average and reaching its destination sooner. This clearly shows that broadband “noisy” stimuli provide better localisation performance in real-world environments. Withington also highlights the importance of ecological validity in testing, stating that “it would be of little value to test siren sounds by asking people seated in a room [...] where they thought the sound originated”.

Walton et al (2022) tested several designed sounds for electric scooter alerts. Their primary goal was to determine which sounds are both effective and not annoying. They tested their sounds using a 16-loudspeaker periphonic (with-height) Ambisonics reproduction system and a simulated “pass-by” of an electric scooter travelling at 20km/h playing the sound under test, with one of two background noise recordings, either a city park with distant traffic or a “road traffic noise scenario” recorded directly next to a road. The pass was simulated as being 2m behind the listener and travelled either left to right or right to left. Participants were asked to identify which direction the scooter was approaching from, and later, rate the sounds under test for annoyance. Reaction time was also recorded. They found that, with no additional alert sounds, detection rate for e-scooters in a 60dBA environment was as low as 23%. It is reasonable to assume bicycles would have similar low detection rates, as they are similarly quiet. They also found that “high frequency, saw wave type sounds” provided the best detection and localisation performance out of any under test.

Catchpole et al (2004) posit that auditory warning signals carry three key pieces of information: “what (semantic), where (location) and when (perceived urgency)”. They performed two left-right localisation tests, both using stimuli played through speakers and recorded using a dummy head ahead of time, which was then played over headphones to participants. The source locations were at 5, 10, and 20 degrees to each side of the median plane. The first experiment determined the ability of listeners to localise six sounds, three of which were tonal alert sounds, and the rest of which were noise signatures. The tonal signatures were an up sweep, a down sweep (both between 1 and 3kHz), and a pure tone (at 2kHz). The noise signatures were broadband white noise (as a control), white noise with a notch from 1-3kHz, and white noise with a notch from 1-10kHz. The test used a two-alternative forced choice paradigm. A second test determined the ability of listeners to both localise a sound and identify which tonal signature it contained. For this test, six sounds were used: the same three tonal alert sounds as the first test, and three compound sounds made of each tonal sound added to the 1-3kHz notched noise from the first test. They found that the noisy sounds were both localised more accurately, and, in most cases, localised quicker. They also found that the notched noise signatures were harder to localise than the full-band white noise, but were still localised correctly more often than the tonal signatures on their own.

Edworthy et al (2023) tested the localisation ability and perceived pleasantness, urgency and other subjective qualities of synthesised sounds for a smart bell project. They tested two groups of 8 sounds, named “polite” and “urgent” by the designer, in a localisation test. An 8-loudspeaker one-in-front system was used, with the subject seated in the centre. For each test, a sound was played, and the subject was asked to select which loudspeaker it came from using a computer program running on a touchscreen computer. Response latency was recorded, with responses being considered incorrect if more than 8 seconds passed. Each sound was played once from each speaker, meaning a total of 64 trials for a “run” of one group. Each subject tested each group twice. The paper does not provide any information on the nature of the sounds beyond the fact that “factors such as pitch, speed, inharmoniousness and so on” were used, meaning it is impossible to draw conclusions from the data about how the acoustic qualities of sounds affect localisation. However, the format of the test is useful to study from, particularly the use of a computer program to collect responses.

Jens Blauert's book Spatial Hearing (1997) serves as an excellent reference work for the fundamentals of spatial hearing and sound localisation, as well as providing a useful preamble on the design of subjective tests such as listening tests. The book discusses inter-aural time and level differences, head-related transfer functions, and their uses in localisation and their consequences, such as the so-called cone of confusion. He noted that cone of confusion errors (eg. front-back reversal) happen more often with short sounds, whereas longer sounds allow the listener to move their head and gather more information about the sound event and therefore are localised better. He also noted that localisation based on interaural time differences is less effective above 800Hz, and entirely ineffective above 1.6Khz.

Wood and Bizley (2015) tested relative sound localisation with respect to signal to noise ratio and differing frequency content of stimuli. In one experiment, subjects were seated in the centre of a partial ring of loudspeakers, and asked to identify which direction a sound had "moved" in the presence of one of three pre-determined background noise levels. For each trial, the background noise levels ramped up over the course of a second, then after a random delay from 50ms to 1050ms, three noise pulses were played from a reference loudspeaker at a rate of 10Hz. After another 130ms, another three noise pulses were played from a target loudspeaker 15 degrees either to the left or right of the reference. The test was a two-alternative forced choice design, meaning participants had to either pick left or right, with no option to say they did not know. After the subject gave a response, the next trial would begin automatically after 1 second. In a second similar experiment, the background noise levels were fixed, and three different stimulus spectra were used: one broadband, one lowpassed at 1kHz, and one bandpassed between 3 and 5kHz. In the first experiment, they found that localisation performance decreases with decreasing signal-to-noise ratio (SNR). In the second experiment, they found that performance was decreased with the lowpassed and bandpassed noise, but was not significantly different between the two.

Investigation was made into the feasibility of using Ambisonics, wavefield synthesis, or binaural reproduction systems for a localisation test. Ambisonics and wavefield synthesis are spatial audio reproduction systems that function by recreating the sound field around the listener's head (Gerzon, 1973; Spors et al, 2008). The precise details of their respective function is beyond the scope of this project. Binaural systems use a collection of head-related impulse responses to "bake" localisation cues into monaural audio streams, which are then reproduced over headphones, with the end result that the same signals arrive at the eardrum and inner ear as would arrive from real sources in physical space. The listener, therefore, perceives the sound as originating from the desired location (Roginska, 2017).

Wierstorf et al (2017) tested localisation accuracy in HOA and wavefield systems using binaural simulations. In a pre-study, they found that a well-designed binaural reproduction system with head tracking has an error similar to the human discrimination threshold, and therefore was valid to use for other localisation tests, even using an HRTF measured from a dummy head. In their full study, they tested localisation accuracy of both plane waves, point sources outside the virtual array, and point sources inside the virtual array ("focused sources"). The sources were at fixed positions, but the listener was "moved" around inside the virtual array between each test, allowing them to test localisation accuracy at various

points within the array. In each trial, a looping train of white noise pulses was played through the virtualisation system into the headphones. The listeners had a keyboard on their knees and a laser pointer mounted to the headphones, pointing forward, to avoid “estimation errors”. The listeners were instructed to turn to face towards the sound source, and then press a key on the keyboard to indicate they had localised the sound and wished to make a response. Each trial began immediately after the listener made a response to the previous trial. They found that localisation accuracy was good in systems with even small numbers of loudspeakers, as long as the loudspeaker spacing was also small. For larger loudspeaker spacings the localisation accuracy decreased.

5.2 Summarised theory of sound localisation

The theory of human sound localisation is well documented, and the fundamentals have been known since almost 150 years ago, when Lord Rayleigh proposed his duplex theory in 1877. The two main qualities used for localisation are inter-aural time differences (ITDs) and inter-aural level differences (ILDs). When a sound is produced away from the median plane, the source is closer to one ear than the other. This means it arrives at the closer ear slightly sooner and slightly louder, leading to time and level differences. Localisation based on ITDs becomes inaccurate when the wavelength of the sound is smaller than the distance between the ears, as it’s hard to tell whether the signal is delayed by a fraction of a wavelength or multiple wavelengths. The oft-stated value for this cutoff is around 1.5kHz to 1.6kHz (Blauert, 1997; Catchpole et al, 2004). Likewise, localisation based on ILDs becomes inaccurate at frequencies lower than around 3kHz (Catchpole et al, 2004) because the sound diffracts more easily around the head, resulting in less level difference than there would be at higher frequencies.

A consequence of using these two qualities for localisation is the so-called cone of confusion. The cone of confusion refers to a set of roughly cone-shaped imaginary surfaces extending outwards from the ears. Every point on one of these surfaces has the same difference in distance to each ear, and as such, any sound located on one of those surfaces produces the same ILD and ITD as any other sound located on the same surface. Additionally, ILDs and ITDs cannot provide any localisation of height, as the elevation of a sound source obviously does not change the relative distance to each ear (Blauert, 1997). A common consequence of the cone of confusion is front-back reversals, where a sound to the front of the listener is perceived as coming from behind.

These problems are resolved using the filtering effects of the head and outer ear (pinna). The shape of the pinna, as well as the shape of the rest of the head, shoulders, and torso, impart unique spectral cues for sounds arriving from given directions. These cues are then processed by the brain to resolve any ambiguities and fill in any missing information about the location of the sound. The overall effect of ILDs, ITDs, and head/pinna filtering is referred to as the head-related transfer function, or HRTF (Blauert, 1997).

The consequence of this is that sounds that fill a larger portion of the frequency spectrum are easier to localise, as there is more for the filtering effects to work with, so to speak (Catchpole et al, 2004; Wood and Bizley, 2015). Speech, for example, is quite broadband,

whereas a bell is extremely narrowband (see figures 6.1.3 through 6.1.6). This should result in speech being far easier to correctly localise than a bell.

HRTFs can be measured and then used for binaural headphone reproduction, which is where virtual sound objects are processed using a measured HRTF to produce a stereo audio stream, which is then played over headphones. The pre-processing of the sounds theoretically imparts the same spectral cues as the listener's head and ears would have done, and so the listener perceives the sound as coming from the intended location. In practice, however, this does not function perfectly. HRTFs are unique per-person, which means for the ideal effect the listener's HRTF must be measured before reproduction. This is less relevant to the topic at hand, but became relevant during the design of the listening test.

Another method of resolving ambiguities is small involuntary head movements (Blauert, 1997). Blauert states that, upon hearing a sound, humans make two involuntary head movements. The first is in any direction, and is merely to gather more cues from which to draw localisation conclusions. The second is after the sound has been localised, and is a movement towards the perceived sound source. Longer sounds, therefore, allow more time for cues to be gathered on either side of the first involuntary movement, increasing localisation accuracy.

Louder sounds are also easier to localise, as necessarily the signal to noise ratio of a louder sound amidst the same background level is higher, and higher signal to noise ratio correlates with better localisation performance (Wood and Bizley, 2015).

6 Methodology

6.1 Design process of study

The obvious question that presents itself is "which type of alert sound is easier to localise", and the obvious way to answer that question is with a localisation test. Several tests within the literature use an Ambisonics or binaural reproduction system to create virtual sound sources (Jin et al, 2002; Walton et al, 2022), but just as many use discrete physical loudspeakers (Withington, 2019; Edworthy et al, 2023). The pros and cons of each were researched in preparation for a test design. Additionally, the format of the test had to be decided. Some tests in the literature were only concerned with left-right localisation (Catchpole et al, 2004) but others positioned sound sources all around the listener (Withington, 2019; Edworthy et al, 2023). It was decided to test source positions on a horizontal plane surrounding the listener, as this represents the locations a bike could come from in a real situation. Elevation was not tested, as it is unlikely for a bike to be approaching a listener from above or below, and even if one was, knowledge of the azimuthal location is more important to avoiding a collision.

Very early in the test design, other ideas were considered. One idea was a field study in which an experimenter would ride a bicycle around a busy or semi-busy area while making use of an alert strategy under test, and noting down the response of pedestrians. This was quickly rejected for obvious safety, practicality, and research ethics concerns. Another idea

was considered, in which listeners would be asked to identify and locate multiple types of event within an acoustic scene, and would only be told after the fact that the test was specifically focused on bicycles. This would potentially reduce the impact of listeners knowing that they were being tested and as such paying more attention to bicycle sounds than they otherwise would. Similar such tests are discussed a little in section 7.2: Limitations and further investigation.

Ambisonics, wavefield synthesis, and binaural reproduction systems were all considered for the listening test, but rejected. They were considered because all three would theoretically allow a sound source to be placed at any position, allowing for much greater resolution of tested source positions.

Wavefield synthesis was rejected partially because the wavefield synthesis system present in the Salford University Acoustics Labs is not comprised of full bandwidth loudspeakers, and because the control system is difficult to use at best (Will Bailey, personal communication, 20th October 2025).

Ambisonics was rejected because an Ambisonics reproduction array with enough loudspeakers for high localisation accuracy could simply be used for a test with discrete loudspeakers instead, with no risk of errors caused by the reproduction system. Additionally, at the time of consideration, it was not known if a viable setup of computers and computer programs talking to each other would be possible with the resources available. In retrospect, this was a wise decision, as the technical problems encountered during the actual test suggest that the use of an Ambisonics system would have been difficult bordering on impossible.

Binaural headphone reproduction was rejected, not for reasons of accuracy, but to avoid the need for HRTF measurements and a potentially complex head-tracking system, and of course any inaccuracies caused by errors in such a setup. Additionally, it was again unknown if such a setup would be possible to assemble with the resources available.

As such, discrete physical loudspeakers were chosen, specifically a circular (or polygonal) array with the listener seated in the centre. 8 loudspeakers were used for ease of setup; a 16 loudspeaker setup was considered but rejected because it would take twice as long to set up and pack down each time. A two-in-front system was used (ie. loudspeakers at 22.5 degrees, 67.5 degrees etc) as in a one-in-front system half of the loudspeakers are directly in a cardinal direction, meaning localisation errors (particularly cone of confusion errors) are unlikely to occur for those sources.

It was decided very early on to write a computer program to run the tests. This both reduces the workload of the experimenter, potentially allowing more tests to be run and reducing any chance for mistakes in data entry; removes any chance of experimenter bias in per-trial stimulus selection; and allows for response latency (and therefore time-to-identify) to be recorded. Time-to-identify data may provide an additional avenue with which to judge “quality” of an alert sound, alongside just ease of localisation. It also somewhat reduces participant workload, as they do not need to translate speaker position to a number or coordinate with the experimenter midway through a test, for example. Additionally it allows the test to proceed at the rate the listener wishes it to, which would not be possible with a

pre-decided sequence of trials rendered into an audio file. This could also be achieved without a computer program by having an experimenter manually start each trial when the listener requests it, but that would increase experimenter workload and possibly be a distraction to the listener or a source of bias. The testing program is explained in detail later.

In retrospect, an 8-speaker setup on one side of the listener could have been used to decrease the angle between loudspeakers and thereby increase the resolution of the test. This approach would have required additional loudspeakers for Ambisonics reproduction, but that would have only brought the total up to 10 or 12. Additionally, it would have required subjects to participate in two or more runs of trials so that both hemispheres were tested.

Four different sounds were selected for the test: voices in the male and female registers of speech saying “excuse me, bike”, and two different bells. Two voices were chosen to reduce the impact of any random features in the way the phrase was spoken for the recording that could make localisation harder, and to find any potential differences in localisability between the two registers. Of note is the “male” and “female” voices were provided by a transgender woman (the author) and a transgender man (pre-hormone replacement therapy) respectively. This likely does not affect any characteristics of the recorded voice.

Two bells were chosen so as to test for differences in localisability between the two. One bell was a “single-impulse” bell, consisting of a sprung hammer that is drawn back and released to strike a metal bell. This style of bell produces a single “ding” when rung. The second bell was a “multi-impulse” bell, consisting of a lever that is pushed to spin an internal mechanism at high speed. Metal weights loosely mounted on the mechanism repeatedly strike an internal protuberance of the bell, repeatedly ringing it. This style of bell is characterised by a more drawn out “dring-dring” sound.

Figure 6.1.1. The two bells recorded for the study, mounted side by side on the handlebar of a bike. On the left is the single-impulse, and on the right is the multi-impulse.

Figure 6.1.2. The internal mechanism of the multi-impulse bell. Also visible is the “bump” on the inside of the body of the bell.

All sounds were recorded at various times in the recording studios in Salford University’s Newton building. An AKG C414 microphone was used for all recordings for its flat frequency response and ease of availability. The microphone’s built-in filter and gain controls were turned off. The following spectrograms were generated using Audacity 3.4.2. Some low-frequency noise is present in these spectrograms, which is believed to be from the air conditioning system in the studios. It was inaudible in the recording and was removed before the test took place. Note that the multi-impulse bell has multiple long-lasting spectral maxima, compared to the single-impulse bell, which only has one. Also note that both voice recordings have much more broadband spectra than either of the bells, which are extremely narrowband outside of the initial impulse or impulses.

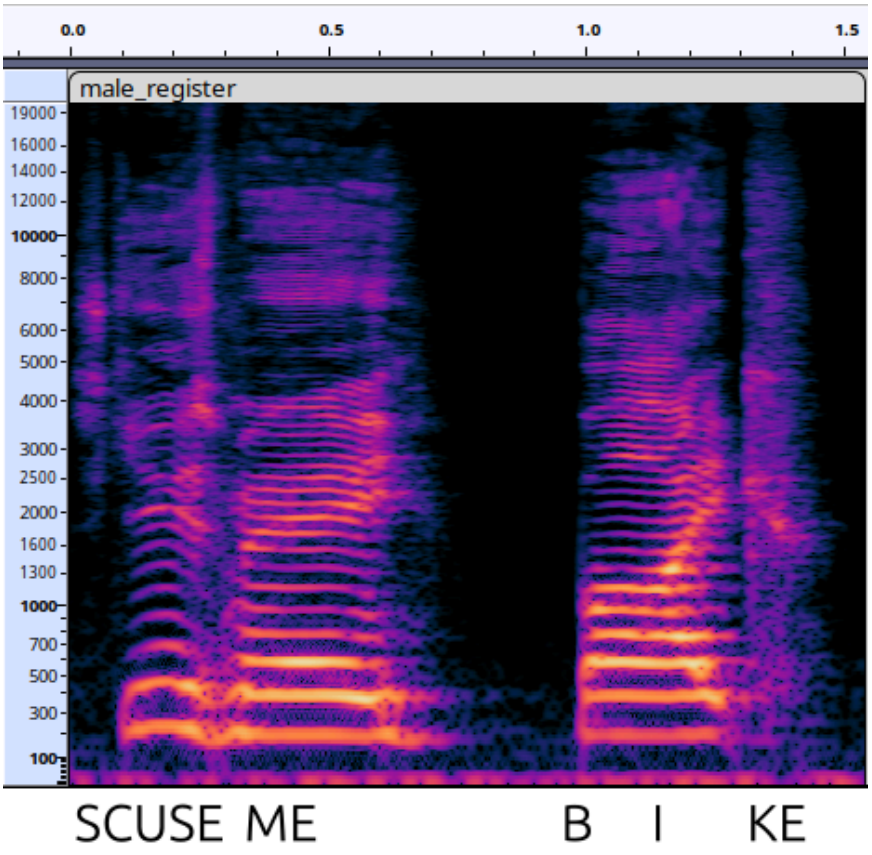


Figure 6.1.3. An annotated spectrogram of the "male register" recording.

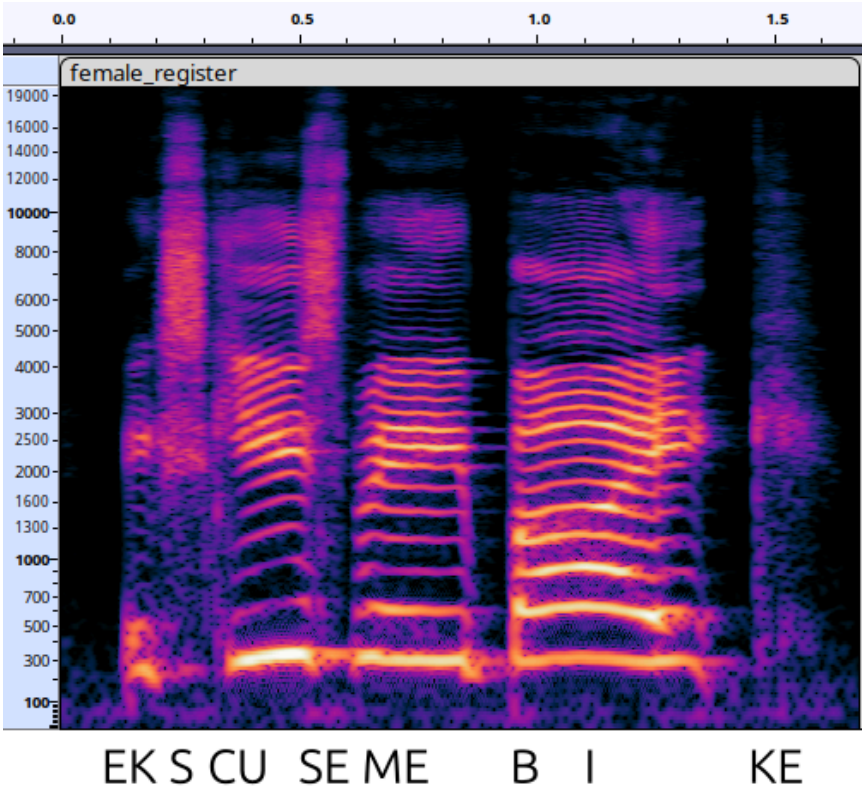


Figure 6.1.4. An annotated spectrum of the "female register" recording.

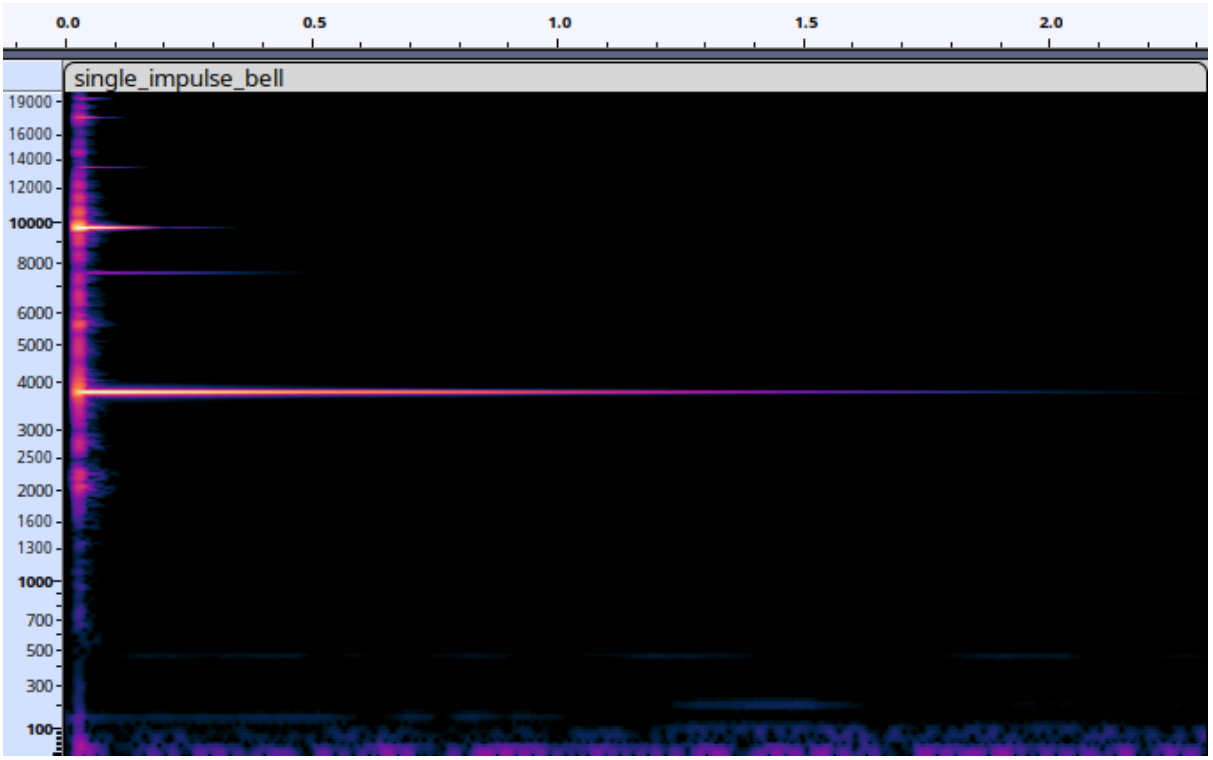


Figure 6.1.5. A spectrogram of the single-impulse bell recording.

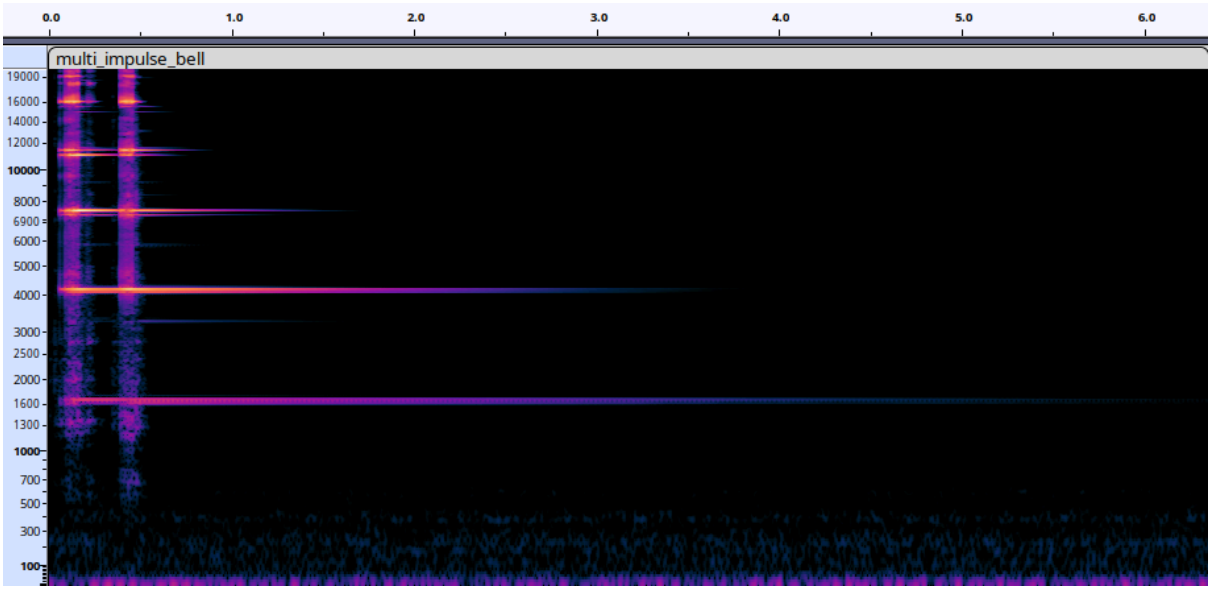


Figure 6.1.6. A spectrogram of the multi-impulse bell recording.

Unlike some other localisation tests found in the literature, which played the stimuli in a continuous loop, ending when the listener made a response (eg. Wood and Bizley, 2015), the test performed in this study only played each stimulus once per trial. This is for ecological validity reasons, and represents a worst case scenario where a cyclist only provides a warning once on approach to a pedestrian.

It was decided that each stimulus would be played twice from each source location, resulting in a total of 64 trials per test. This value was chosen to provide a good balance between data generated per participant and test duration. It was expected that a shorter test would encourage more participation, as each participant would have to give up less of their time.

The presence of background noise was planned from very early on, likewise for ecological validity reasons. Additionally, the pilot test made evident another reason for its presence. Decreasing the signal to noise ratio of the sounds under test would theoretically increase the amount of mistakes made (Wood and Bizley, 2015), making it easier to perform statistical analysis on the produced data. An Ambisonics recording from the Eigenscape project was used. Ambisonics reproduction was used because the 8-loudspeaker array used for the localisation test could easily work a dual purpose as an Ambisonics reproduction array, and Ambisonics produces a reasonably realistic recreation of the recorded environment. Eigenscape is a dataset of acoustic scenes principally intended for machine learning research, available for free in first- and third-order Ambisonics (Green and Murphy, 2017). Originally, it was planned for at least two background noise recordings to be used, one of a quiet pedestrian area and a second of an area with car traffic, however due to time constraints, the test was only run with a recording from a pedestrian area. The specific background noise used was Pedestrian Zone 1. The first-order Eigenscape Lite version was chosen partially because the reproduction system was only an 8-channel pantaphonic (without height) system, and partially to reduce the size of the download. The entire Eigenscape Lite dataset fits within a single 12.6GB zip archive, which is smaller than any of the individual location recording sets from the full third-order dataset. It was decoded using SPARTA AmbiDEC running in Ardour 8. The Pedestrian Zone recording was chosen as it represents a likely situation in which a pedestrian would encounter a bike.

It was hypothesised that the voice recordings would be easier to localise as they have a much wider bandwidth and maintain roughly the same amplitude throughout their duration, both of which should contribute to easier localisation (Catchpole et al, 2004; Blauert, 1997). This is compared to the much more narrowband bell recordings that drop off in level quite quickly. However, bells are louder by ~6dB, meaning they would be masked less by background noise, potentially increasing localisation ability again (Wood and Bizley, 2015).

6.2 Pilot test

A simple left-right localisation test was performed some time before the full-scale test took place. The purpose of this test was partially to get some impression of the localisation performance of the sounds tested, and partially as a trial run to see what perhaps unexpected things would need to be considered for the full scale test, such as how necessary background noise would be.

The test took place in a performance space in the Newton Building studios at Salford University. Sources at 5, 10 and 20 degrees from the median plane were used, located two metres away from the participant at roughly head height. Genelec 1029A active loudspeakers were used. Only two loudspeakers were used, which had to be moved between each run of trials. This was partially due to space constraints in the performance space, partially to reduce setup/pack down time, and partially so that the test could be run from a 2-channel audio interface connected to the experimenter’s laptop. The loudspeakers were calibrated using a pink noise source inside Ardour 8 and a Svantek 971 sound pressure level meter. Only the male register voice and single-impulse bell were used for the pilot test, both to keep the test simple and because the multi-impulse bell sound was yet to

be decided on and recorded. Each combination of sound and location was played four times, giving a total of 16 trials per angle. The order of the trials was randomised separately for each angle, then assembled into pre-made sound files in Avid Pro Tools. Each sound was prefixed with a spoken number to avoid participants losing count of which trial they were on. The sounds were played back using Ardour 8 and an M-Audio MobilePre 2-channel USB audio interface, with the level of the sounds calibrated to approximately the same level as a real voice and real bell coming from the same location. Participants were asked to write down on paper which side they thought the sound had come from after each trial.

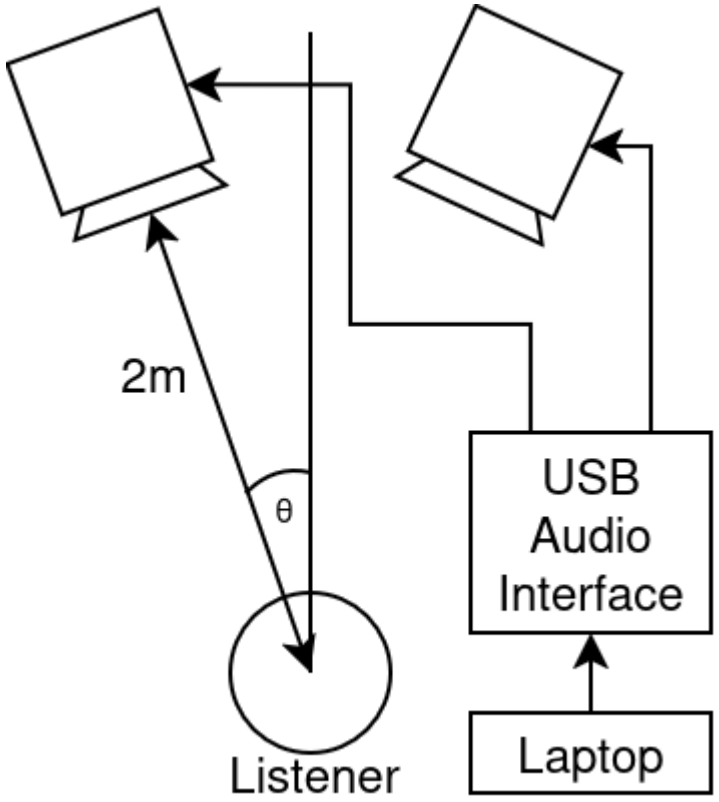


Figure 6.2.1. Diagram of the pilot test setup (not to scale). Theta is the angle from the loudspeakers to the median plane, and is either 5, 10 or 20 degrees.

Two volunteers took part in the pilot test, both self-reported normal listeners, meaning a total of 96 trials were performed. In total, only four mistakes occurred. All mistakes were made when the speakers were 5 degrees from the centreline, and they were evenly split between the voice and the bell.

These results obviously did not provide much use in and of themselves. However, they indicated that simply scaling up the pilot test to an 8-loudspeaker circular array and playing sounds at test participants would not provide much if any useful information, as not enough mistakes would be made to make meaningful comparisons between sounds. This highlighted the importance of adding background noise to the full-scale test to reduce the signal-to-noise ratio and increase the number of mistakes made for both sounds, as, if only a small amount of mistakes are made, it becomes impossible to tell if the difference in the number of mistakes made between groups is truly reflective of an underlying phenomenon, or merely an artefact of chance.

The pilot test also affirmed the utility of a tester program. Entering even only 96 datapoints into a computer by hand to compare them against the correct locations took a while, and required plenty of double- or even triple-checking. As discussed later, the full scale test produced well over five times as many datapoints, so entering them all by hand would have been untenable.

6.3 Tester program

The tester program was written in Python. It displays a simple GUI made using Tkinter, Python's built-in bindings library to the Tk graphics framework, which consists of a large central button labeled "start" and a set of 8 numbered buttons surrounding it, in the same layout as the loudspeaker array. When the program is run, it creates a randomly shuffled list of all the trials set to occur, ie. two instances of each combination of each sound and source location. The start button is used to start each trial. When the button is pressed, the program waits one second, then fetches the next trial from the list and plays the specified sound from the specified loudspeaker. When the user presses a surrounding numbered button, the program records the user's response, moves to the next trial, and then waits for the start button to be pressed again. The sound, correct location, the user's response, the time the sound was played and the time of the user's response (both in seconds since the Unix epoch) are all printed out to the terminal in a comma-separated format. When in actual use, the output of the program is "piped" into a file to be recorded.

The program plays sounds by sending MIDI messages to a separate program (or even a separate computer). MIDI (Musical Instrument Digital Interface) is a simple but flexible digital messaging protocol, used for controlling audio hardware or software such as synthesisers, samplers, or mixing desks. It allows for simple messages to be sent from one program or computer to another, such as "play this note" or "change X parameter to Y value" (MIDI Association, 1996). The utility of this for the task at hand is obvious. On the other end of the MIDI connection from the testing program, there are 8 copies of a digital sampler, each on a separate MIDI channel and each corresponding to one loudspeaker. Each sampler has all four sounds under test loaded, calibrated to the desired level, and set to play when certain notes are sent by the testing program. This allows the testing program to easily request playback of any sound from any loudspeaker. An earlier version of the software used one sampler and sent MIDI control change messages to simple gain plugins to select which loudspeaker the played sounds were sent to. This turned out to require a lot more effort while setting up the Carla patch than originally thought, and additionally once it became apparent the sounds would have to be played on a Mac running Pro Tools (see section 6.4: Technical restrictions), the software was changed to the "stupid but simple" approach used for the actual test. Under the original plan, the sampler in use would be LSP Sampler, a free and open source software sampler, loaded in Carla, a free and open source generic plugin host and audio/MIDI patchbay. However, due to several technical restrictions (discussed later), the actual test used instances of Native Instruments Battery 4 loaded in Avid Pro Tools.

The source code of the testing program is available in Appendix D.

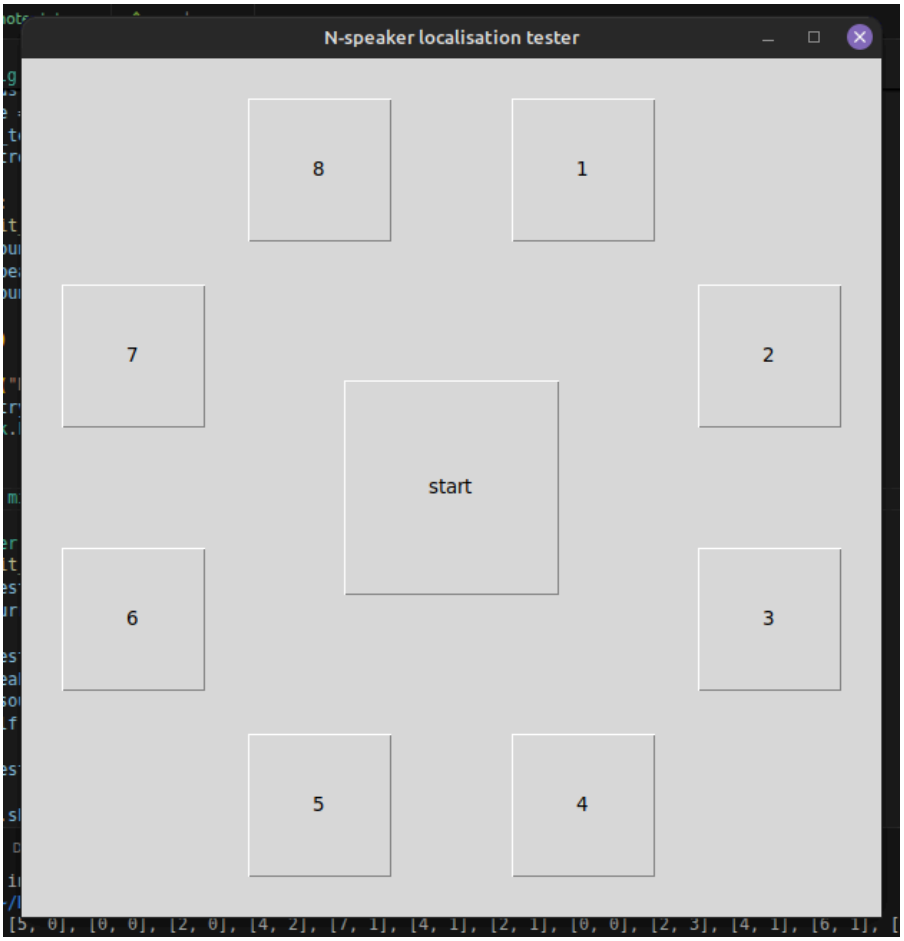


Figure 6.3.1. A screenshot of the tester program's GUI

6.4 Technical restrictions

Originally, the full scale test was planned to take place in the Listening Room at Salford University's Acoustics Labs. This location offers a wide open space where a loudspeaker array could be set up and a listener positioned in the middle. However, due to myriad technical difficulties, the test actually took place much later than originally planned in the control room of a recording studio on campus.

The first technical hurdle encountered was the fact that none of the 8-channel audio interfaces available for student use supported the Linux operating system, which the experimenter's laptop runs. More specifically, no drivers had been published by their manufacturers. With this in mind, alternative plans were considered. First was the usage of a Max/MSP patch instead of a Python program and surrounding software. Such a patch was in fact created, using the js object to control the playback of sounds and recording of responses, but it was soon discovered that neither the laptops available for student use (which run Windows and as such *can* talk to the audio interfaces available) or the Macs present in the studio control rooms had Max/MSP installed at the time of the test.

The next plan considered (and the one that ended up being used) was to connect the experimenter's laptop and the studio Mac via MIDI. The use of RTP-MIDI (AKA AppleMIDI)

was considered. RTP-MIDI is a network protocol that allows MIDI connections to be made over a standard network connection (Arduino, 2026). This was quickly rejected, as while RTP-MIDI drivers and software for Linux does exist, none of it is particularly easy to configure, and additionally it was assumed the network set up at the university would not allow such a connection.

The solution, then, was to purchase two USB-MIDI converter devices and a MIDI gender changer to physically connect the two devices. This proved unreliable in practice, likely because the USB-MIDI converters were so cheap, and so the actual test setup utilised the MIDI Merge functionality of the Studiologic SL88 Studio keyboard in the control room. The keyboard has a hardware MIDI input port, and the MIDI Merge functionality allows all incoming messages on that port to be forwarded to the keyboard's MIDI output.

This allowed MIDI to be sent out of the laptop via a USB-MIDI converter, into the keyboard, and from there into the Mac and the eight instances of Native Instruments Battery 4 (a commercial drum sampler, and the only sampler with the required functionality available on the studio Mac) hosted in Pro Tools. Each instance of the sampler was on a separate instrument track, with the MIDI input filtered to its respective channel. The output of the samplers, alongside the pre-decoded Ambisonics recording, was sent out of the computer into the mixing desk present in the studio (an AWS 948 Delta) and from there to the loudspeakers, which were each placed on an insert send from the desk.

The main consequence of these technical problems is that the test ran several weeks later than planned, meaning the potential participant pool was much smaller. A secondary consequence was the test taking place in a less-than-ideal listening space. There was a large mixing desk positioned in between the listener and the rear two loudspeakers, as well as a computer monitor on an equipment rack roughly between loudspeakers 3 and 4. These obstructions may have reduced localisation acuity to the rear and rear left of the listener, however, because they would have an identical effect on all sounds under test, the decision was made to go ahead with a non-ideal test rather than waste more time in the pursuit of perfection.

Another source of delays was the availability of facilities. The tests were eventually run over the university's easter break, as this was the only time that long periods of studio time were available to book.

6.5 Actual test

The physical setup for the test was as follows: 8 Genelec 1029A active loudspeakers were positioned equidistant around the circumference of a circle of radius 1.8m, in a two-in-front configuration. They were positioned using a laser spirit level with an angle measuring tool. They were then calibrated using a pink noise source from inside Pro Tools and a Svantek 971 sound pressure level meter, which itself had been calibrated with a B&K Type 4231 calibrator. The levels produced by the bicycle bells and voices had previously been recorded at the same distance of 1.8m (80dBA and 74dBA respectively), and so the output level of the samplers could be adjusted to match. The samplers were then attenuated by 15dB, to simulate a source approximately 10m away. The background noise was adjusted to a level

of 55dBA at the listener position. While undertaking the test, listeners were seated in the centre of the loudspeaker array with a laptop on their lap, running the testing program. The interface was controlled with the trackpad built into the laptop. Listeners were briefed on the nature of the test and the interface of the testing program, then left to complete the test while the experimenter left the room (so as to reduce distractions or possible bias). Each test consisted of 64 separate trials: two instances of each combination of sound and source location, presented in a random order. With the test complete, they were thanked and sent on their way.

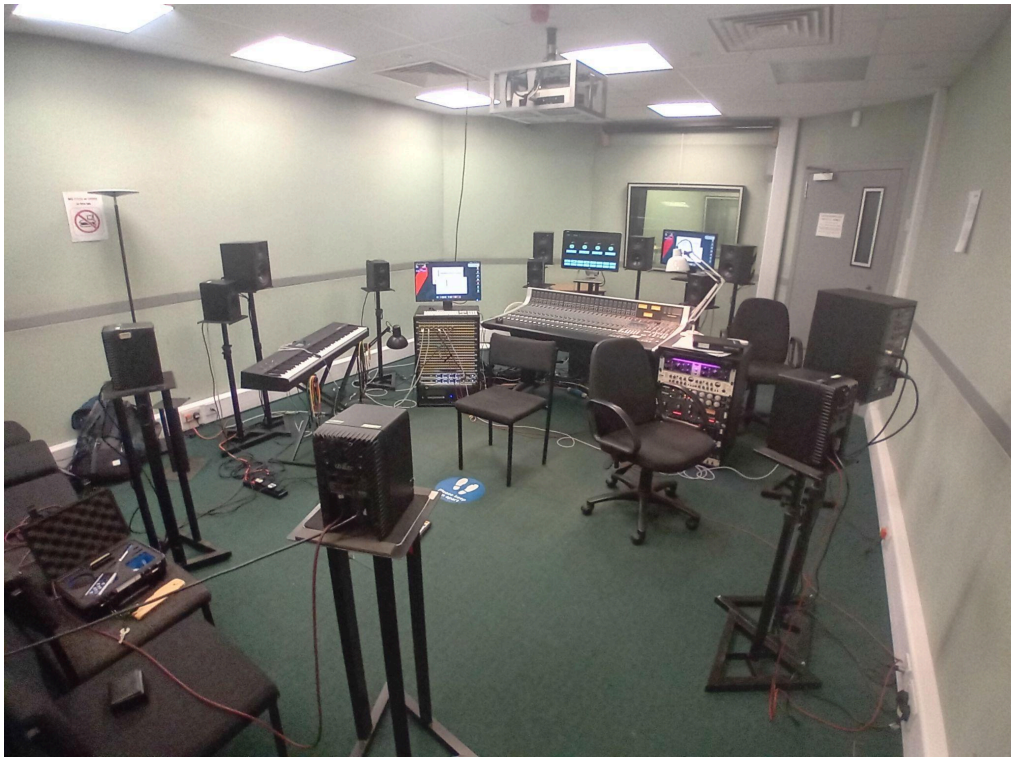


Figure 6.5.1. Testing setup. The laptop is hidden behind the nearest speaker.

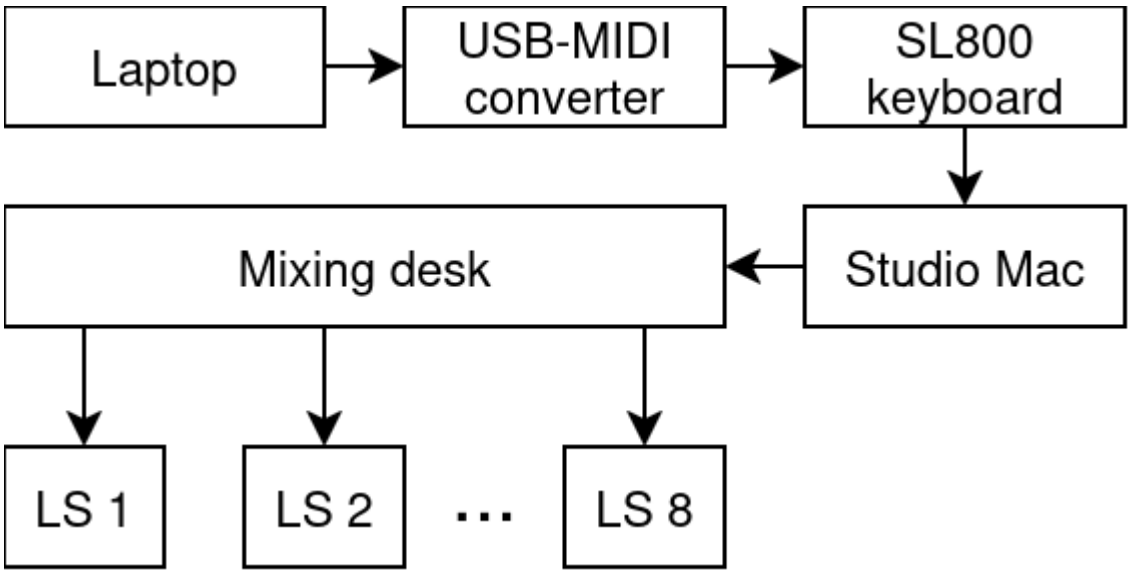


Figure 6.5.2. Signal flow diagram of the testing setup.

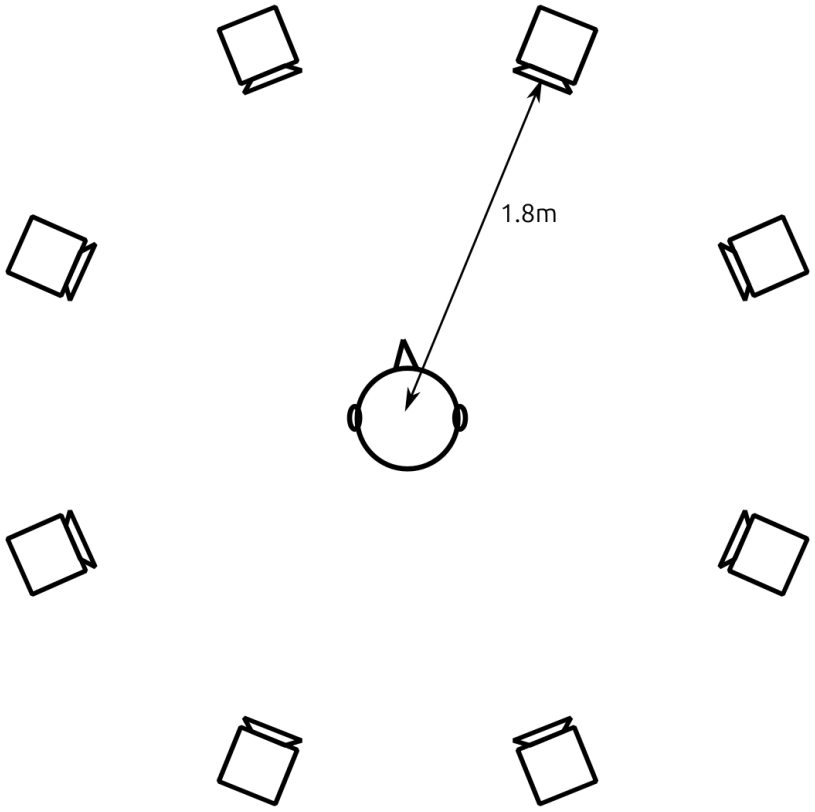


Figure 6.5.3. Diagram of speaker positions relative to listener.

7 Results

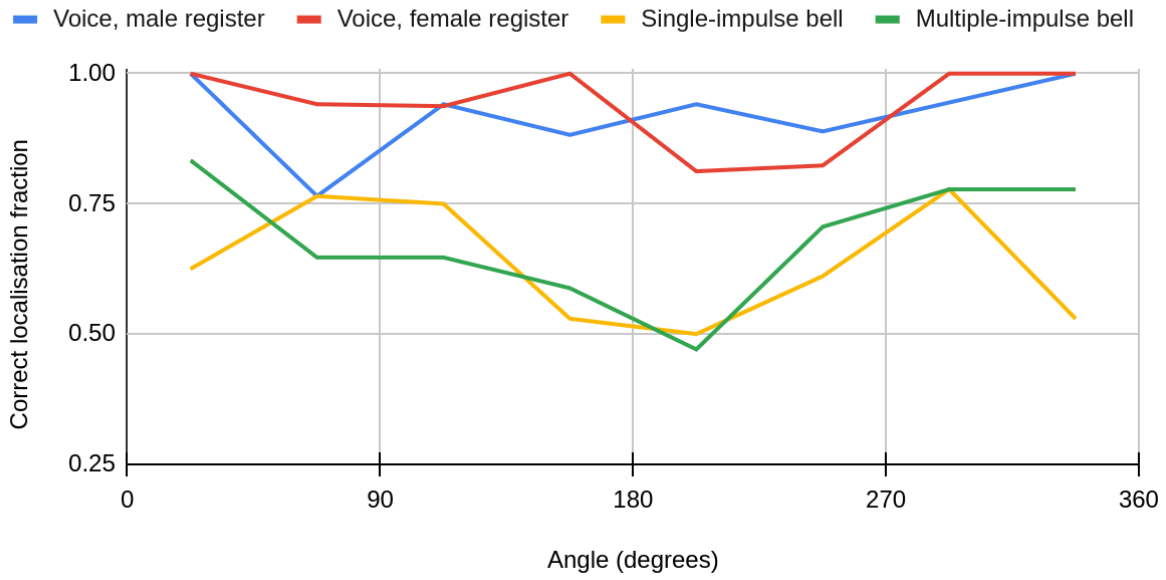
Nine listeners took part in the test. All were self-reported normal listeners, with no hearing problems or impairments. One test was accidentally ended early (just over halfway through) due to the listener incorrectly interpreting the test as finished. As such, 548 individual trials were recorded, roughly evenly distributed between each sound.

Several data processing operations throughout the analysis pipeline, such as sorting, were performed with a collection of ad-hoc programs written in Rust.

Of the trials, 272 were with a voice and 276 were with a bell. 253 of the voice trials ended in a correct localisation, with 19 incorrect; whereas 182 of the bell trials were correct, with 94 incorrect. A chi-squared test was performed, and showed that these two groups are independent ($p = 0$). However, further chi-squared tests showed that the differences in correct localisation frequency between the two different voices and between the two different bells were not statistically significant ($p = 0.287$ and $p = 0.222$ respectively). This shows that voices are significantly easier to localise than bells, but there is no significant difference in localisation ability between the vocal registers tested or between the bells tested.

Additionally, the correct response fraction with respect to source location was calculated and graphed (fig. 7.1). No statistical tests were performed on this as it is mostly a curiosity compared to the main focus of the study. In the graph, both bells show a clear decrease in localisation accuracy at the rear, while the voices do not. This may be indicative of some underlying phenomenon, or it may simply be caused by the small sample size and the low error rate for the voices. Additionally, it may be due to the non-ideal test environment, as there were several pieces of equipment directly behind listeners, possibly making localisation harder.

Fig 7.1. Fraction of correct localisation judgements at each angle



Some participants stated subjective opinions on which sounds seemed easier to localise. Two stated that the voices were easier to localise, while two stated that the bells were easier, one of which saying that the multi-impulse bell specifically was easier.

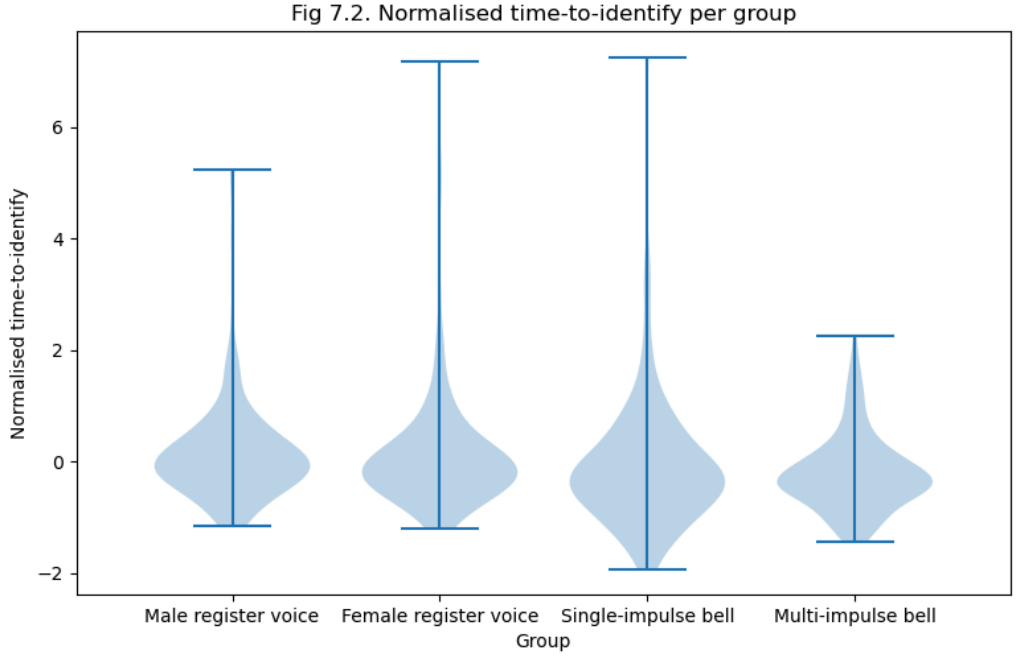
Time-to-identify data was also analysed. The response latency for each trial was calculated from the “sound played” and “response received" timestamps output by the testing program. Then, because each sound clip under test had a different sized delay between the start of the clip and the onset of the sound, the calculated response latency was adjusted to account for this delay. These adjusted latencies will henceforth be referred to as “raw adjusted response latency”. The response latencies then z-score normalised per listener, such that each listener’s mean response latency was zero, and each listener’s standard deviation was 1. Z-score normalisation was performed using a program written in Rust using the meansd library, and was intended to counteract any participants being generally slower than others, and also to remove the time taken for listeners to input their response after they had made a localisation judgement. It was assumed that each participant would take some amount of time after making a localisation judgement to move the mouse on the computer and click a button to respond, and that that length of time would be roughly constant per-listener. The z-score normalised latencies will henceforth be referred to as “normalised response latency”.

Of note is that the response latency value does not represent a true “reaction time”. It is possible or even likely that listeners were aware that a sound had been played long before they made and input a localisation judgement. The experimenter performed several trial runs of the test at various times in the setup process, and definitely experienced this happening.

A simple analysis of means showed that both bells had lower normalised response latencies on average, with the multi-impulse bell being the lower of the two. However, this on its own does not prove anything, so further analysis was performed.

A one-way ANOVA was performed between the four groups of normalised response latencies (each group being all the tests performed for a given sound) using the XLMiner Analysis ToolPak for Google Sheets. It was found that the underlying mean values are different for at least one group ($p = 0.00354$). Post-hoc Tukey-Kramer tests were then performed. The Q critical value was calculated for each group, and then the largest value was used ($Q_{crit} = 0.31$), to be conservative. It was found that both vocal registers were significantly different to the multiple-impulse bell, but no other comparison was found to be significant.

Fig 7.2. Normalised time-to-identify per group



7.1 Discussion of results

This study was intended to determine or at least provide some insight on whether voices are easier to localise than bicycle bells in real world environments, and, more generally, which sound type is “better” as an audible alert for a bicycle. It is important to keep in mind that the actual results collected only reflect the performance of voices and bells in a simulated and controlled listening test environment. In the test, the voice recordings used showed significantly better localisation accuracy than the bell recordings used. However, the

multiple-impulse bell recording used showed significantly lower time-to-identify than both of the voice recordings used, despite its lower localisation accuracy.

The increased localisation accuracy of the voice recordings is in agreement with conventional wisdom found in the literature. The wider spectral bandwidth of a voice and the fact that the voice recordings used maintain roughly the same amplitude throughout their duration both contribute to easier localisation (Catchpole et al, 2004; Blauert, 1997), compared to the much more narrowband bell recordings that drop off in level quite quickly. The comparatively lower time-to-identify exhibited by the bell recordings is potentially due to their higher peak amplitude (approximately 6dB higher than the voice recordings) and the fact that they reach this peak amplitude very quickly after the onset of the sound, and possibly due to their high frequency content. Walton et al (2022) found that higher pitched alert sounds resulted in lower reaction times.

Interestingly, listeners did not seem to universally perceive the bells as being harder to localise. Of the four that expressed subjective opinions on ease of localisation, two stated that the bells were easier to localise, with one of the two specifically saying the multi-impulse bell seemed easier to localise. The reason for this is unknown.

7.2 Limitations and further investigation

The UI of the testing program was non-ideal. It provided almost no feedback as to the current state of the test, and only informed the user that the test was complete by printing a message to a terminal window in the background. For example, if a user clicked the central “start test” button while a trial was currently ongoing (ie. the program was waiting for a response) nothing would happen, with no indication given that nothing was going to happen, so to speak. As previously mentioned this caused confusion on at least one occasion. Three very large outliers were found in the raw adjusted response latency data, indicating some difficulties with usage of the program. Additionally, a touchscreen could have been used to improve ease of use and hopefully reduce random variation in the time-to-identify data, as in Edworthy et al (2023).

An obvious point of improvement is that more tests could have been performed. Due to the automated nature of the test and almost all the data processing, this would not have increased experimenter workload by much, and so would have been very feasible had the test not been delayed and had the volunteer pool been larger. As a substitute for additional participants, each test could have been longer, with 4 instances of each possible trial rather than two. This naturally would have doubled the length of the test, and during the actual testing period it was decided to use 2 instances of each trial to keep the test short and hopefully increase participation.

The time-to-identify data exhibited lots of overlap between groups, with many outliers. Even though the statistical analysis found several parts of it significant, this and the known issues with the testing program interface call it into doubt somewhat. A subsequent study could use a more fluid response entry system to reduce the impact of delay caused by response entry and hopefully reduce noise in the data, as well as potentially reduce the complexity of the localisation task to a left-right or front-back task to focus specifically on time-to-identify or

reaction time, as in Catchpole et al (2004) or Walton et al (2022). Such a response entry system could be a similar one to the test performed in this study, but with a touchscreen for input rather than a laptop trackpad, or a physical unit built with physical buttons, as in Withington (2019).

The listening environment no doubt reduces the quality of the data collected, as well as does not match the expected environment to encounter a bicycle in. In daily life, one is likely to encounter a bike in an open or semi-open urban space, but the test was performed in an acoustically treated studio control room, with several pieces of equipment within the listening area. This likely does not change any differences between the results of the different sounds under test, but the observations made in a more appropriate acoustic environment would likely be different.

Unlike in Walton et al (2022), the noise produced by the bike by itself as it moves was not present in the test, and all sounds under test were reproduced at static positions. Had the test been performed using an Ambisonics or binaural reproduction system, source positions could have been moved in real time to better simulate the experience of being approached by a bicycle. However, this would necessitate a change in how localisation responses were recorded, as recording the perceived location of a moving sound obviously requires more nuance than the same would for a static sound.

Another potentially confounding issue is that listeners naturally knew they were being tested, and as such were actively listening out for the sounds under test. This means that the data collected and conclusions synthesised by the study only reflect ease of localisation of sounds in a controlled listening test environment, rather than the more general and more useful information of how effective sounds are at alerting to the presence of a bike. Very early in the design of the study, a test was considered where listeners would be asked to identify or locate many different types of sound within a virtual environment, without being told the test was focused on bicycle alerts. This would potentially provide a more realistic look into the efficacy of sounds at alerting an initially unaware listener. Additionally, the listeners could be instructed to perform tasks such as solving simple mathematical problems, sorting objects, or perhaps navigating an environment to see how good the sounds under test are at alerting a listener who isn’t listening, so to speak.

This study also did not concern itself with how appropriate certain alert sounds are for the task at hand. Everybody knows that a bell means a bike is nearby, but, as we hear voices all the time, a shouted voice does not necessarily have the same association. This ties into the points discussed above about the inherent bias present in tests.

A few test participants expressed subjective opinions on the ease of localisation of the sounds under test, which raises the possibility of running a subjective survey alongside the more objective localisation accuracy results. Such a survey could also include annoyance or pleasantness ratings, as in Walton et al (2022) or Edworthy et al (2023). It would be interesting to study perceived ease localisation in comparison to actual (ie. statistically measured, as in this study) ease of localisation.

8 Conclusion

In this study, a localisation test was carried out on four recordings of two broad types of acoustic alert strategies for bicycles - shouted voices and bicycle bells. Time-to-identify (the time taken from the onset of a sound to a listener making a localisation judgement) was also recorded and analysed. In the test, and in agreement with the theory, voices showed significantly higher localisation accuracy than bells, suggesting that, by certain criteria, a voice constitutes a better alert sound. However, one of the bells under test (a multiple-impulse bell with a rotary internal mechanism that strikes the body of the bell multiple times) showed significantly lower time-to-identify, despite its lower localisation accuracy.

This study also hopefully demonstrates the benefits of using a computer program to orchestrate listening tests, namely those of reducing experimenter and participant workload and allowing more types of data to be recorded.

9 References

Arduino. (February 2026). *AppleMIDI Documentation.* https://docs.arduino.cc/libraries/applemidi/

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Catchpole, K.R., McKeown, J.D. and Withington, D.J. (2004). Localizable auditory warning pulses. *Ergonomics*, *47*(7), pp.748-771.

Cycling UK. *Cycling UK's cycling statistics.* https://www.cyclinguk.org/campaigns/statistics

Department for Transport. (2025, 22 October). *The Highway Code - Rules for cyclists.* https://www.gov.uk/guidance/the-highway-code/rules-for-cyclists-59-to-82

Edworthy, J., Brown, R. and Wessel, C. (2023, September). Where's that bike? Sounds and metrics for a smart bicycle bell. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 67, No. 1, pp. 2519-2524). SAGE Publications.

Gerzon, M.A. (1973). Periphony: With-height sound reproduction. *J. Audio Eng. Soc*, *21*(1), pp.2-10.

Green, M.C. and Murphy, D. (2017). EigenScape: A database of spatial acoustic scene recordings. *Applied sciences*, *7*(11), p.1204.

Jin, C., Best, V., Carlile, S., Baer, T. and Moore, B. (2002, April). Speech localization. In *Audio Engineering Society Convention 112*. Audio Engineering Society.

MIDI Association. (1996). *MIDI 1.0 Detailed Specification*. https://midi.org/midi-1-0-detailed-specification

Roginska, A. (2017). Binaural audio through headphones. In *Immersive sound* (pp. 88-123). Routledge.

Spors, S., Rabenstein, R. and Ahrens, J. (2008), May. The theory of wave field synthesis revisited. In *124th AES convention* (pp. 17-20). Audio Engineering Society (AES).

The Pedal Bicycles (Safety) Regulations 2010

Vaillancourt, V., Nélisse, H., Laroche, C., Giguère, C., Boutin, J. and Laferrière, P. (2013). Comparison of sound propagation and perception of three types of backup alarms with regards to worker safety. *Noise and Health*, *15*(67), pp.420-436.

Walton, T., Torija, A.J. and Elliott, A.S. (2022). Development of electric scooter alerting sounds using psychoacoustical metrics. *Applied Acoustics*, *201*, p.109136.

Withington, D.J. (2019). Localisable alarms. In *Human factors in auditory warnings* (pp. 33-40). Routledge.

Wood, K.C. and Bizley, J.K. (2015). Relative sound localisation abilities in human listeners. *The Journal of the Acoustical Society of America*, *138*(2), pp.674-686.

Appendices

Appendix A: Project log book

Weeks 1-4 (September and October): initial research, initial ideation on test design

10th October: project plan submitted. Research started in earnest

October-December: Literature review, pilot test design, full test design, various inquiries into equipment and facilities available.

3 November 2025: Recording and level measurement of male register voice and single-impulse bell.

1 December 2025: Recording and level measurement of female register voice

12 December 2025: Pilot test.

December-January: Pilot test analysis, subsequent amendments to full test design, creation of poster presentation.

1st February 2026: What was thought to be a multi-impulse bell was ordered, but it turned out to be a single-impulse with a sideways lever.

4th February 2026: Presented poster presentation.

February: Full test design finalised and tester program written.

25th February 2026: Multi-impulse bell delivered, recorded and measured.

The experiment was delayed compared to the initial plan because I had to investigate what options I had available of where to perform it and what to perform it with, including trying to find audio interfaces that worked.

The Listening Room was booked for the 12th March, and then it was found that neither the audio interface plan nor the Max/MSP plan were going to work, so the booking was cancelled and alternative options were considered.

15th March: USB-MIDI converter dongles delivered. Several studio sessions were booked over the next week or so to attempt to debug and trial run the setup.

24th March: First working test setup. However, no tests were run on this day because I ran out of time after debugging the setup.

7-8th April: First actual testing setup. The system was built on the afternoon of the 7th, some tests were ran, then it was left in place overnight to continue on the 8th

10th April: Second testing setup. The system was built in the morning, tests were ran, and then it was disassembled in the afternoon.

Remainder of April: Data analysis and writeup, including generation of all graphs, spectrograms, etc.

Appendix B: Project plan

The following is the project plan submitted in October before the study began in earnest.

Ease of localisation of acoustic warning strategies for bicycles

Student: Allison Edwards
Supervisor: Trevor Cox

When cyclists and pedestrians meet, it is important that pedestrians are able to identify not just the existence but also the location of an approaching cyclist. Anecdotal evidence suggests that the traditional acoustic warning device used - a bicycle bell - has less than ideal localisation characteristics, with human speech being a better option. This project will identify the difference in efficacy between the two, if any such difference exists.

Experiment

Participants will be asked to identify the direction of an incoming bike in a virtual audio environment, created with either a ring of loudspeakers or the wavefield synthesis system in the listening room. Accuracy and time-to-identify will be recorded for each trial, in order to compare the differences between the two alert sounds being studied. Additionally, background noise will be played to more accurately simulate the real world situation under test.

Objectives

The following objectives are considered required for the project:

- Perform a literature review, to gain an understanding of the mechanics of sound localisation and what characteristics of sounds can make it easy or hard, and additionally to find out how other similar studies have been performed
- Analyse recordings of various bike bells and voices to make some more confident predictions
- Investigate practicality of using the wavefield synthesis system for the experiment
- Investigate input methods for participant answers
- Record or otherwise obtain bells, voices and environmental noise for experiment
- Run the experiment with at least two acoustic environments: quiet pedestrianised area (eg. a university campus) and an area with car traffic
- Perform statistical analysis on test results

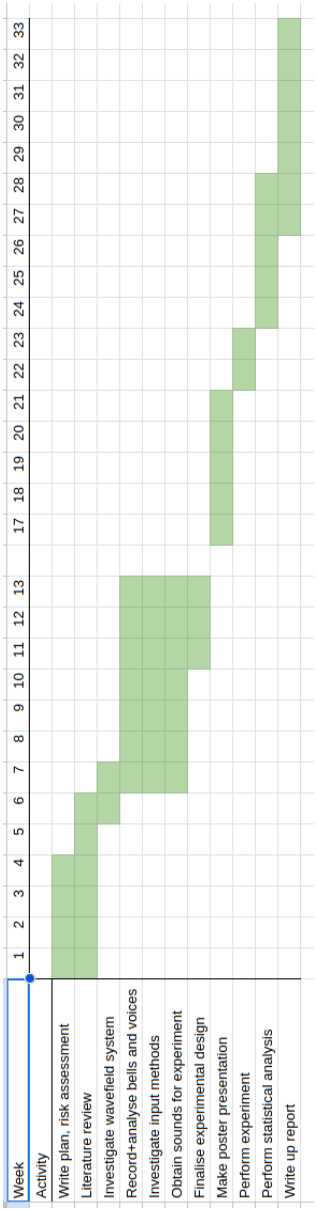
The following objectives are not required, but would be interesting to complete if time permits:

- Test different bells in different frequency registers, and test the difference between male and female voices, if any
- Test additional acoustic environments (eg. areas with trams, areas with lots of wind, reverberant spaces such as subways)
- Test additional non-bell acoustic warning devices such as horns

Equipment and space requirements

The experiment will require access to either a large number of loudspeakers and the audio booth, or the listening room and the wavefield synthesis system, for potentially several hours spread out over multiple blocks.

Signature of student: Allison Edwards Date: 10/10/2025
Signature of supervisor:
Signature of second reader:



Appendix C: Project risk assessment

The following is the risk assessment submitted alongside the project plan.

RISK ASSESSMENT FOR RESEARCH AND PROJECTS
COVER SHEET.

Student's surname	Edwards
Student's forenames	Allison
Section of School	SEE
Degree	BSc Sound Engineering and Production
Title of Project	**Ease of localisation of acoustic warning strategies for bicycles**
Brief description of the project	A localisation test will be performed to determine if there is any difference in localisation ability between bycicle bells and the human voice, and if so, what that difference is.
Supervisor's name	Trevor Cox
Reserve supervisor in case of absence.	
Locations where work will be carried out	

It is the responsibility of the PI / Supervisor to ensure the risk assessment is suitable and sufficient and that risk assessments are reviewed should any significant changes to the project / research be made.

Risk assessment (RA) tracking;				
Hazard	**RA required**	**RA complete**	**Risk level as identified in RA**	**Authorised as suitable and sufficient by supervisor**
Initial Risk Assessment.	yes	yes	1	
Physical / mechanical / electrical / animal	yes	yes	1	
Biological – microorganisms	N/A			
Chemical	N/A			
Radiation	N/A			
Travel / Fieldwork	N/A			
Host Organisation	N/A			

I shall be working in the establishment named below (where applicable):
Name; Address; Tel;

Emergency Contact Details.		
Role	**Name**	**Telephone & email**
Student		
Academic Supervisor		
Host organisation Supervisor (where applicable)		

Please complete the following risk assessment forms as dictated by the proposed project.

Guidance on how to complete the assessments can be found in the Project RA handbook.

To save paper please delete the sections that are irrelevant to your project.

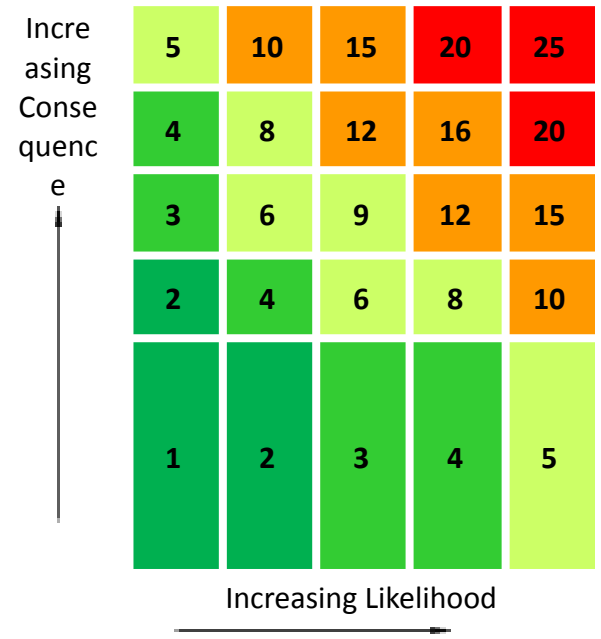
Contents.

Initial Risk Assessment.

Location:		Task/Activity/Environment:		Date of Assessment:
Identify Hazards which could cause harm:		Identify risks = what could go wrong if hazards cause harm:		
No.	Hazard	No.	Risk	
	Physical / mechanical / electrical / animal		See section	
	Biological			
	Chemical			
	Radiation / Lasers			
	Lone working			
	Travel / Fieldwork			
	Disposal of waste material			
List groups of people who could be affected:			What numbers of people are involved?	
What existing precautions are in place to reduce risks?				Risk level with existing precautions
What additional actions are required to ensure precautions are implemented/effective or to reduce the risk further?				Risk level with additional precautions

Is health surveillance required? YES/NO	If YES, please detail:	
Who will be responsible for implementing the precautions:		By When:

Risk Rating:



17-25 Unacceptable – Stop activity and make immediate improvements/seek further advice
10-16 Tolerable – look to improve within specified timescale
5-9 Adequate – Look to improve at next review
1-4 Acceptable - No further action, but ensure controls are maintained

Guide to using the risk rating table:

Consequences	**Likelihood**
1 Insignificant – no injury	**1 Very unlikely** – 1 in a million chance of it happening
2 Minor – minor injuries	**2 Unlikely** – 1 in 100,000 chance of it happening
3 Moderate – up to three days absence	**3 Fairly likely** – 1 in 10,000 chance of it happening
4 Major – more than three days absence	**4 Likely** – 1 in 1,000 chance of it happening
5 Catastrophic – death or disabling	**5 Very likely** – 1 in 100 chance of it happening

Physical / Mechanical / Electrical / Biological (Animals) Risk Assessment.

Location: Newton building - listening room and/or audio booth	Task/Activity/Environment: Localisation test	Date of Assessment: 10/10/2025

Identify Hazards which could cause harm:		Identify risks = what could go wrong if hazards cause harm:	
No.	Hazard	No.	Risk
	Exposure to high sound levels		Hearing damage
	Trip hazards (cables etc)		Minor physical injury

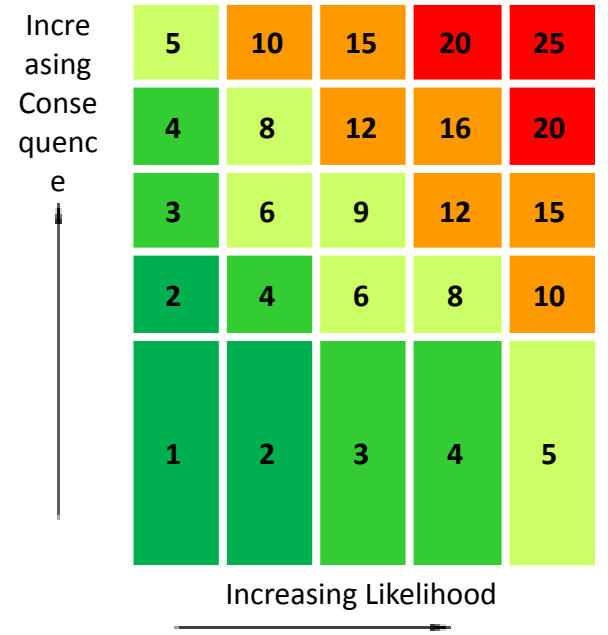
List groups of people who could be affected: Test participants	What numbers of people are involved?	~15, but only one at a time

What existing precautions are in place to reduce risks?		Risk level with existing precautions
No.		
	Measure sound level inside test setup to ensure it never reaches dangerous levels	1
	Tape cables to the floor or otherwise make sure theyre out of the way	1

What additional actions are required to ensure precautions are implemented/effective or to reduce the risk further?		Risk level with additional precautions
No.		

Is health surveillance required? YES/NO	If YES, please detail:	
Who will be responsible for implementing precautions:		By When:

Risk Rating:



17-25 Unacceptable – Stop activity and make immediate improvements/seek further advice

10-16 Tolerable – look to improve within specified timescale

5-9 Adequate – Look to improve at next review

1-4 Acceptable - No further action, but ensure controls are maintained

Guide to using the risk rating table:

Consequences	**Likelihood**
1 Insignificant – no injury	**1 Very unlikely** – 1 in a million chance of it happening
2 Minor – minor injuries	**2 Unlikely** – 1 in 100,000 chance of it happening
3 Moderate – up to three days absence	**3 Fairly likely** – 1 in 10,000 chance of it happening
4 Major – more than three days absence	**4 Likely** – 1 in 1,000 chance of it happening
5 Catastrophic – death or disabling	**5 Very likely** – 1 in 100 chance of it happening

Appendix D: Tester program code

The following is the source code listing of the Python tester program. It can easily be adapted to other loudspeaker configurations and sounds by editing the static values in the `Config` class.

The source code can also be found at https://github.com/ambyshframber/localisation_tester.

```
import mido

from tkinter import *
from tkinter import ttk

import math
import random
import time
import sys

class Config:
 num_speakers = 8
 is_2if = True
 num_sounds = 4
 base_note = 36
 num_each_test = 2
 base_controller = 12

class Test:
 def __init__(self, sound_idx, speaker_idx):
   self.sound_idx = sound_idx
   self.speaker_idx = speaker_idx
   self.sound_time = None

root = Tk()

root.title("N-speaker localisation tester")
root.geometry("600x600")
frame = ttk.Frame(root).grid(column=0, row=0, sticky=(N, W, E, S))


midi_out = mido.open_output(mido.get_output_names()[1])

class Tester:
 def __init__(self):
   self.test_num = 0
   self.current_test = None

   self.tests = []
   for speaker in range(Config.num_speakers):
     for sound in range(Config.num_sounds):
       self.tests.append([speaker, sound])
```

```
    self.tests = self.tests * Config.num_each_test

    random.shuffle(self.tests)
    print(self.tests)

  def start_test(self):
    if self.current_test:
      return
    if self.test_num == len(self.tests):
      print("test complete!", file = sys.stderr)
      return
    test = self.tests[self.test_num]
    speaker = test[0]
    sound = test[1]

    self.current_test = Test(
      sound,
      speaker
    )
    root.after(1000, self.play_test)

  def play_test(self):
    note = Config.base_note + self.current_test.sound_idx
    msg = mido.Message("note_on", note = note, channel =
            self.current_test.speaker_idx)
    midi_out.send(msg)
    self.current_test.sound_time = time.time()

  def end_test(self, guess):
    if self.current_test == None:
      return
    if self.current_test.sound_time == None: # check if test has been played yet
      return
    data = [self.current_test.speaker_idx, self.current_test.sound_idx, guess,
            self.current_test.sound_time, time.time()]
    print(",".join([str(x) for x in data]))
    self.current_test = None
    self.test_num += 1

tester = Tester()

def on_press(n):
  tester.end_test(n)

def make_button(parent, idx):
  angle_per = 360 / Config.num_speakers
  base_angle = 0
  if Config.is_2if:
    base_angle = angle_per / 2
```

```
 angle = base_angle + (angle_per * idx)
 angle = math.radians(angle)
 x = (math.sin(angle) * 0.4) + 0.5
 y = (math.cos(angle) * -0.4) + 0.5
 ttk.Button(parent, text=idx + 1, command=lambda :
           on_press(idx)).place(anchor="center", relx=x, rely=y, height=100,
           width=100)

for x in range(Config.num_speakers):
 make_button(frame, x)

ttk.Button(frame, text="start", command=tester.start_test).place(anchor="center",
           relx=0.5, rely=0.5, height=150, width=150)

root.mainloop()

print("exiting")
msg = mido.Message("control_change", control = 123) # all notes off
midi_out.send(msg)
```