# Listening loudspeakers: a quieter and more intelligible world

Iain McGregor
Online Sound Design MSc Programme Leader, Edinburgh Napier University



Apple Homepod https://www.apple.com/uk/homepod/

Loudspeakers can now truly be described as ubiquitous, being part of many portable and fixed technologies. They are an important part of everyday life, providing both entertainment and essential information. Unfortunately, growth in utilisation has often created a crowded and noisy acoustic environment. With the increased adoption of voice controlled assistants the concept of having a device that listens to you is becoming more readily accepted. Monitoring the output of a loudspeaker from a listener's perspective can be an effective way for not only improving clarity, but also for minimising the impact that the device has on anyone else sharing the environment. Instigating an intelligent feedback loop replicates what a considerate person does in everyday practice: they listen until it is their turn to speak and then alter their delivery so that it matches more closely what is appropriate for the situation. If someone comes to the door or a telephone rings during a conversation they cease talking, if something noisy occurs that is unimportant they will either talk over it, or stay quiet until the disturbance passes, at which point they repeat whatever might have been missed. If a person is talking quietly they will usually also speak quietly, if a more normal level is expected then it will be matched, and if shouting is required, then a decision can be made whether or not to engage.

Many devices with loudspeakers also have microphones, and these can be used to alter the output in real-time through digital signal processing (DSP). Microphones can be very inexpensive, especially if piezo technologies are utilised, and even in small devices they are relatively easy to incorporate in a discrete manner. The simplest change to achieve using DSP is the volume. The ambient level of sound can be measured and the volume turned up or down accordingly. This means that the output could be set to always be a specified level above the background. It would be relatively simple for the device itself to learn what the desired level is in a noisy environment.




By tracking the changes a user made to the volume during regular use, and then comparing the information with the corresponding background levels, the optimal difference could be calculated. If this leads to excessive peaks, then a form of compression can be applied so that as the sources becomes louder the differences between the loudest and average levels are reduced. Correspondingly, if an environment has a low level of background noise then the signal could be expanded so that there is an even greater difference between the loudest and quietest parts of the audio content.

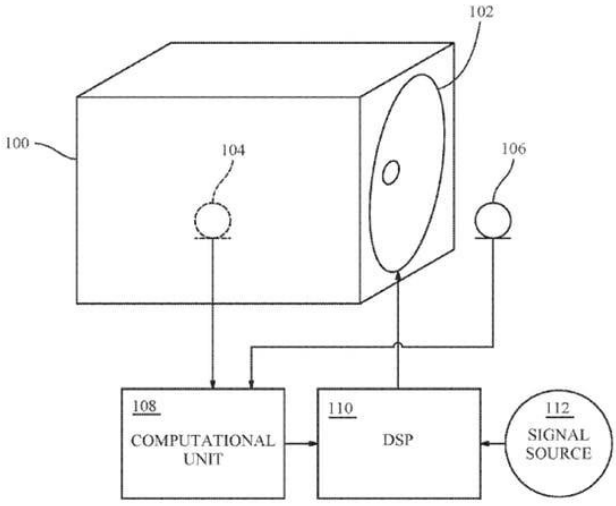
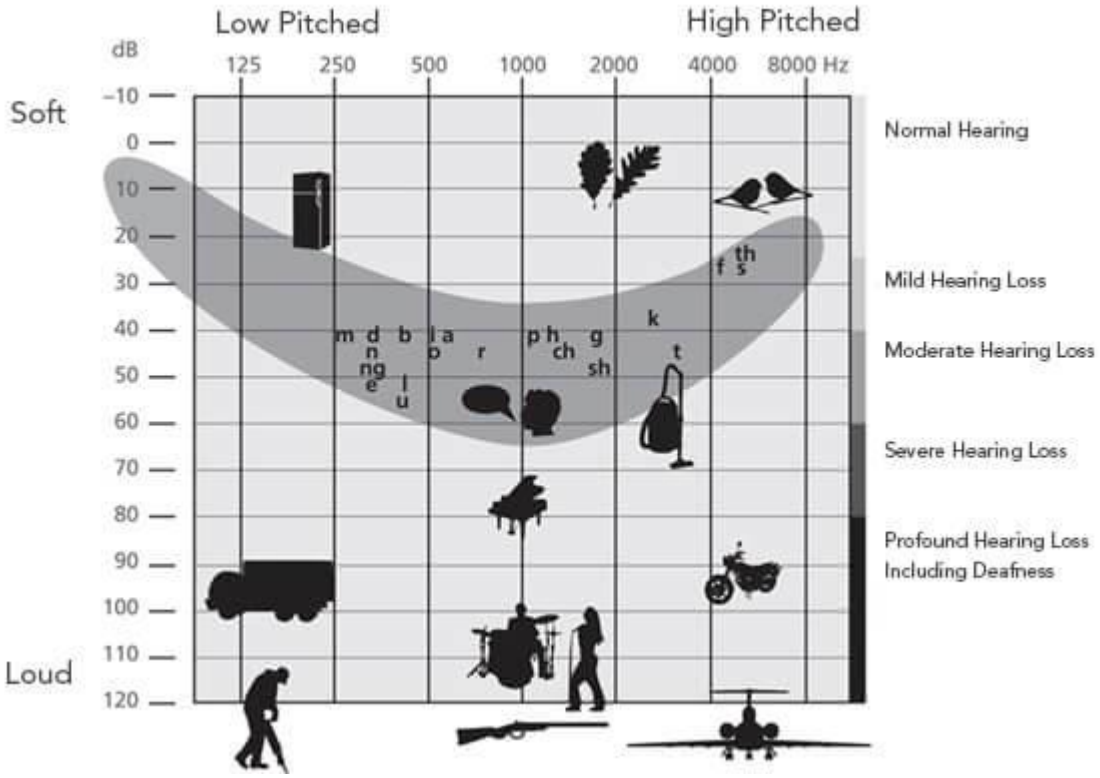
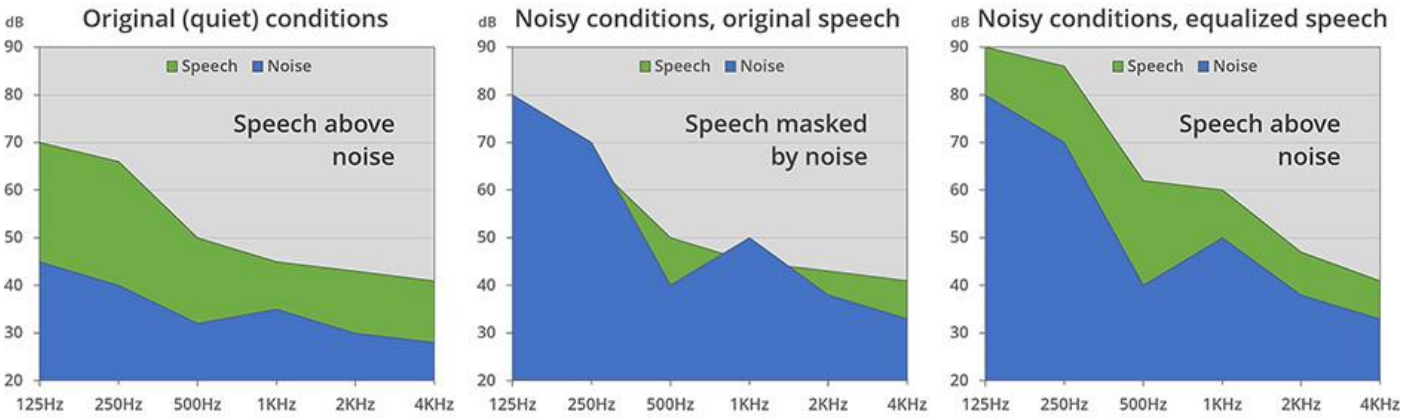



*FIG. 1*

The frequency content of a loudspeaker is already a standard parameter that can be affected in real-time through DSP. Initially designed to correct for shortcomings in room acoustics, the capabilities can be extended to compensate for the effect of perceived equal loudness. In general terms, quieter sounds are experienced as having less bass content than louder sounds, although all frequencies are affected. By linking a microphone at the listener's position to the transmitting device it is not only possible to partially correct the timbre for the acoustics, but also to compensate for the level of listening. By altering the frequency content with volume, the loudspeaker can be perceived as sounding the same no matter the volume it is set to. Any portable device could potentially be enlisted for this purpose, and this could take place in any environment.




Frequencies could also be adjusted according to a person's hearing abilities. If someone cannot hear above a certain frequency a low pass filter could be applied so that redundant frequencies are removed. If a listener has hearing notches (losses) then frequency ranges which are less damaged are typically perceived as louder, and these could be turned down so that clarity is improved. A traditional approach is to make everything louder, which can lead to further hearing loss, as well as cause disturbance to others in neighbouring environments.




Another issue that often leads to excessive volume is masking, where a louder sound source makes it difficult to hear the desired signal. For broadband (wide frequency ranges) the only current solution is to turn the volume up. However, if this is not practical then just like a person repeats something if it is clear that the recipient could not hear it, then an active system could pause output until a temporary masker has subsided. At which point the content could then resume from just before the interruption. If only a narrow range of frequencies are affected then the offending frequencies could be boosted to overcome the issue until the noise subsides. The choice about how to proceed could be established from previous end –user behaviour, if content has been rewound in the past then a pause would be adopted. If the end-user turned the volume up and then down, a temporary increase in the masked frequencies would be applied. Both approaches mean that any adjustment is only for the duration of the interrupting noise.

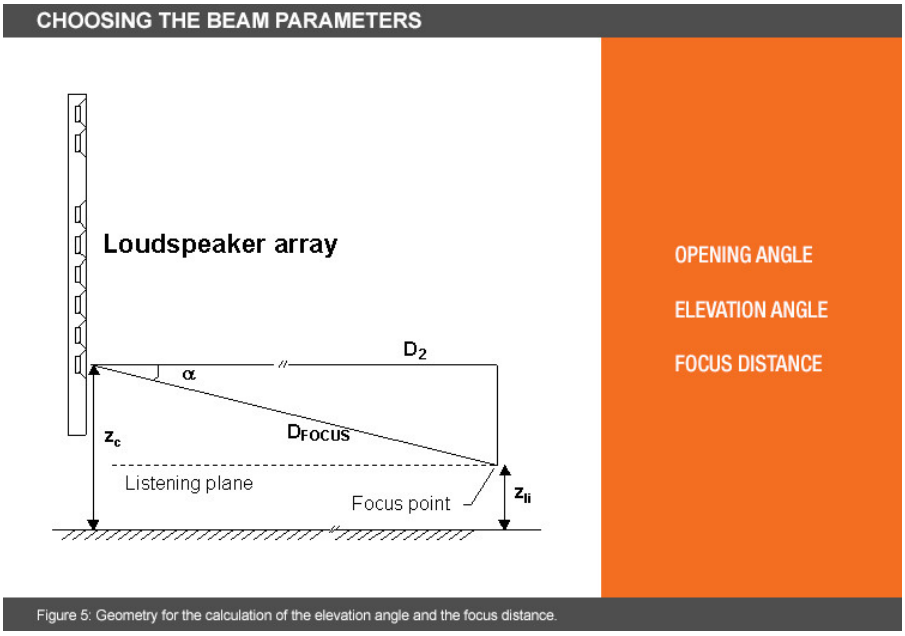
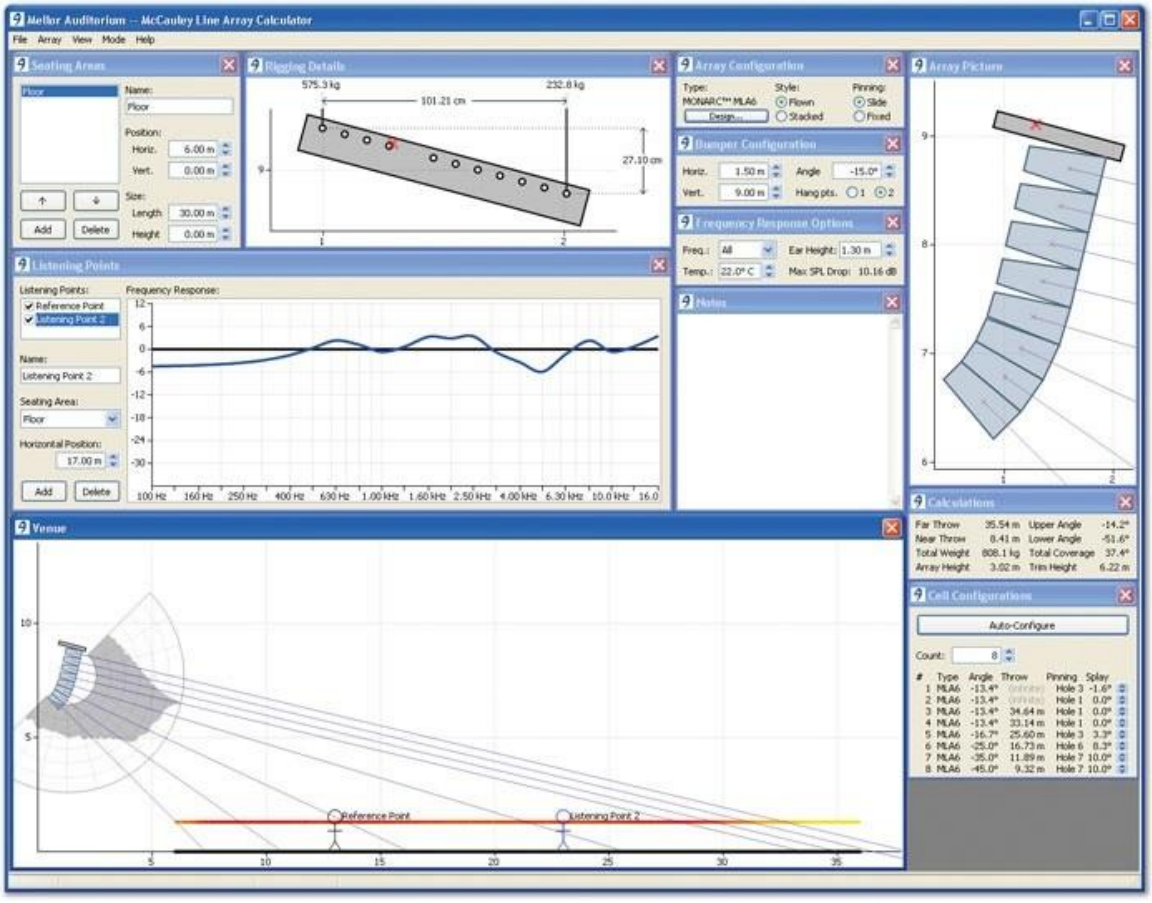
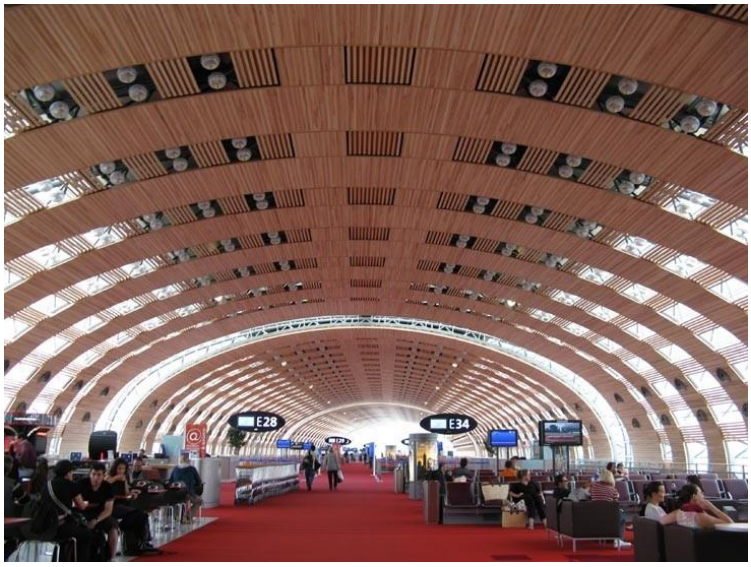



Figure 5: Geometry for the calculation of the elevation angle and the focus distance.

When column array loudspeakers with multiple transducers or drivers are utilised (such as commonly found in places of worship), it is possible to use DSP to spatially align the output without moving the housing. The range of the frequencies that can be steered is dependent on the position and number of drivers as well as the length of the cabinet. Generally, the longer the cabinet the lower the frequency that can be affected. The upper frequency limit is affected by how closely the centres of the high frequency (HF) drivers are to each other. The smaller the distance the higher the frequency that can be affected. These highly specialist loudspeakers were originally designed for environments with high levels of reflections so that the sound can be directed more precisely, and improve intelligibility. By independently altering the timing, frequency content and amplitude of each signal to successive drivers the angle and width of the sound transmission can be altered in real-time through a technique commonly called beam steering.



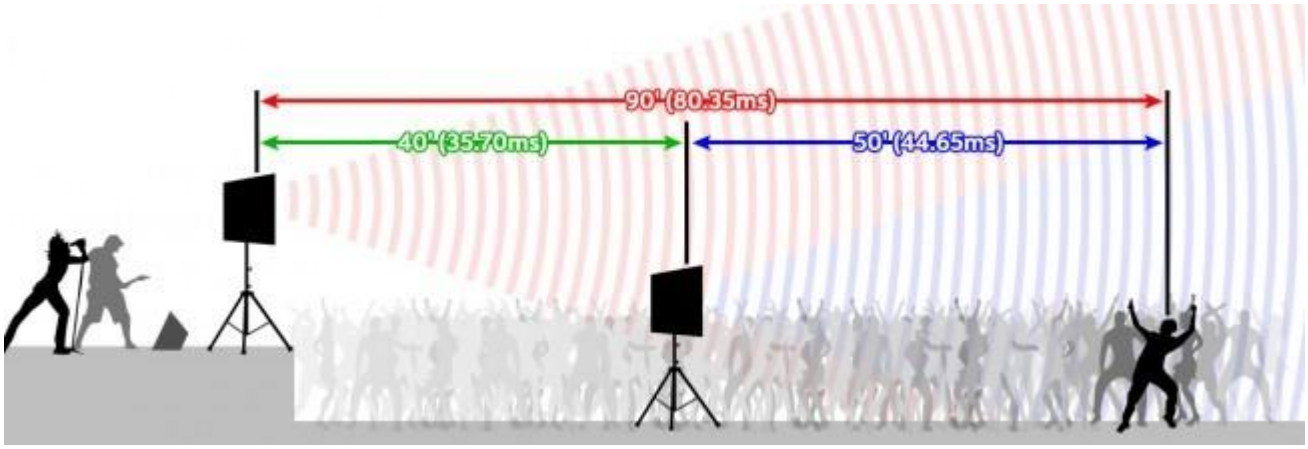
The line array approach, where multiple drivers are stacked above each other already has the natural ability to transmit sound efficiently over long distances. The addition of DSP means that with remote monitoring through a microphone the relationship between direct and indirect sound arriving at the desired listening positions can be altered in real-time. Direct sound represents the waves travelling directly from the loudspeaker to the listener. Indirect sound refers to all of the reflections (often termed reverb or reverberation), which can greatly impede clarity, but which are often used to naturally amplify a sound. Generally, the greater the percentage of reflections compared to the direct source the slower the sound needs to be in order to remain intelligible. This means that in highly reflective environments fast sounds like speech are typically difficult to fully comprehend.



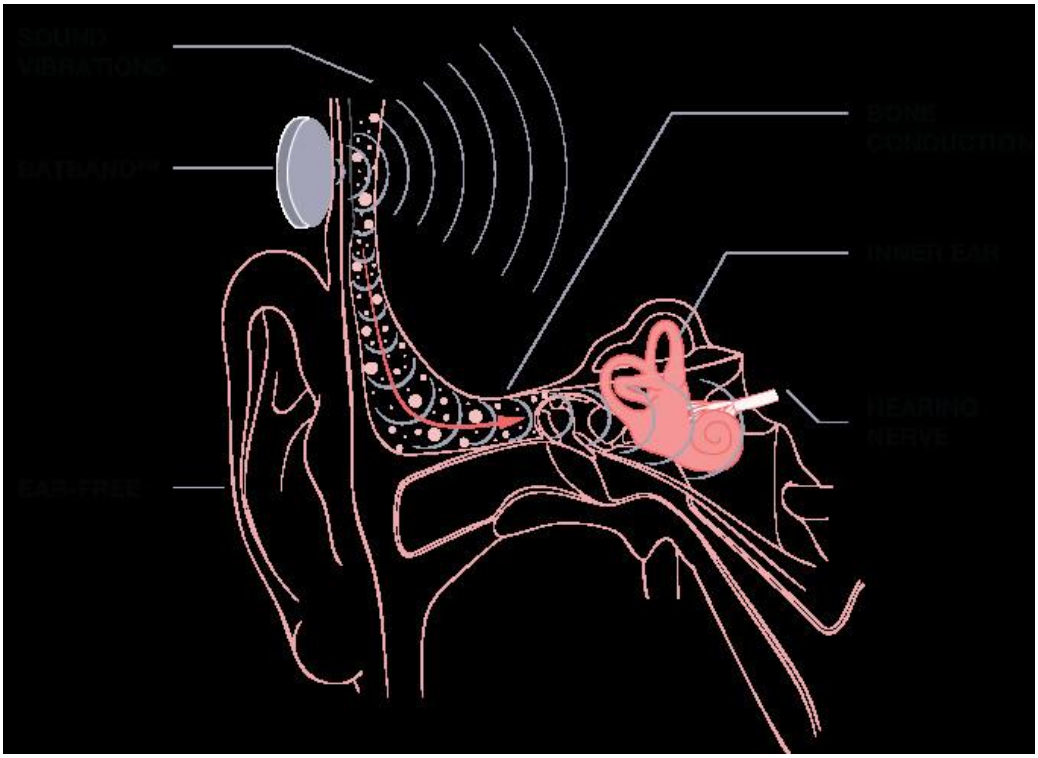

This situation becomes worse when listeners move around within an environment, such as in a train station or airport, where obstruction, occlusion and exclusion become factors. Obstructions mean that only indirect sound can reach the listener but they can still be relatively clear, some of the direct sound, especially the lower frequencies will pass around the object so will typically sound slightly muffled. Occlusions occur when sound has to pass through a barrier like a wall so that both the direct and indirect sound is muffled. The extent of the filtering is entirely dependent on the material the sound is passing through. Exclusions, such as doorways mean that only the indirect sound is spectrally altered, again typically becoming muffled. When these issues are combined with the output from multiple loudspeakers spread across a complex physical environment it becomes obvious why announcements are often difficult to comprehend.



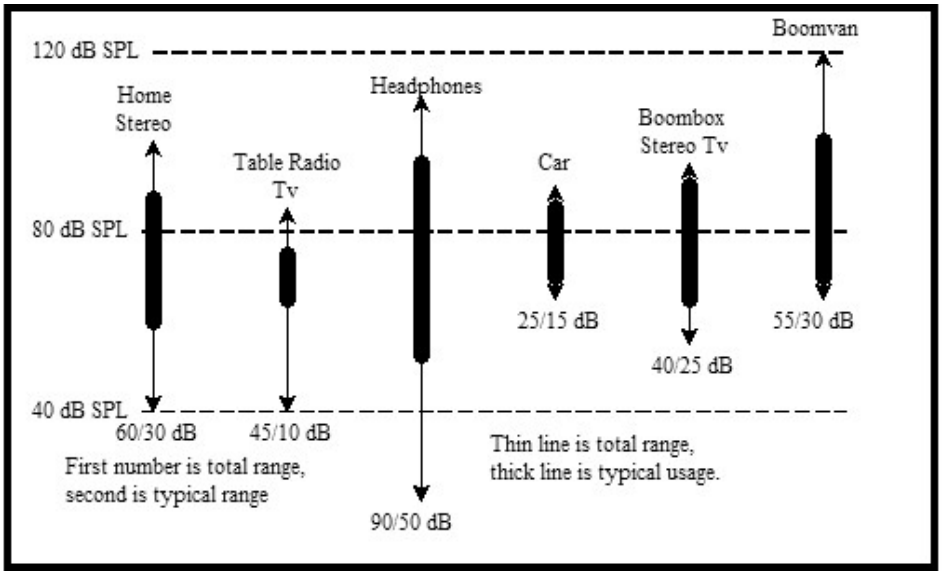
The first and most obvious solution is to check if anyone is there, if not then loudspeakers can be switched off. Microphones can be set to capture the sound of footsteps (short, percussive sounds) and then through triangulation of pitch and amplitude establish which direction they are walking in. The Doppler effect is a well understood principle, a pitch shift upwards and increase in amplitude denotes movement towards an object. Correspondingly, a downwards pitch and decrease in volume communicates movement away from an object. Other cues such as establishing whether a sound is temporarily obstructed, can indicate that a structure or structures has moved in front of a source. Depending on the absorption pattern it could be people, or a vehicle. This will obviously change according to the number of people and the effect on reverberation times has been studied so that it can be used as a method of estimating the population of an environment.




A second solution is time alignment, where attention can be drawn to the intended perceived source of a sound by ensuring that the direct sound arrives first from the desired location. Any sound that is reproduced through additional loudspeakers can be delayed so that it arrives as close as possible in phase with the original signal, which also helps with clarity, and minimising the amount of amplification required. This technique can also be utilised with headphones, or across multiple distinct devices. Sound can be encoded (watermarked) to include inaudible content that can be used to time align it, amongst other functions. This can mean that audio can be streamed across multiple devices and then time aligned internally by the reproduction device so that it is almost perfectly in phase at a known point.



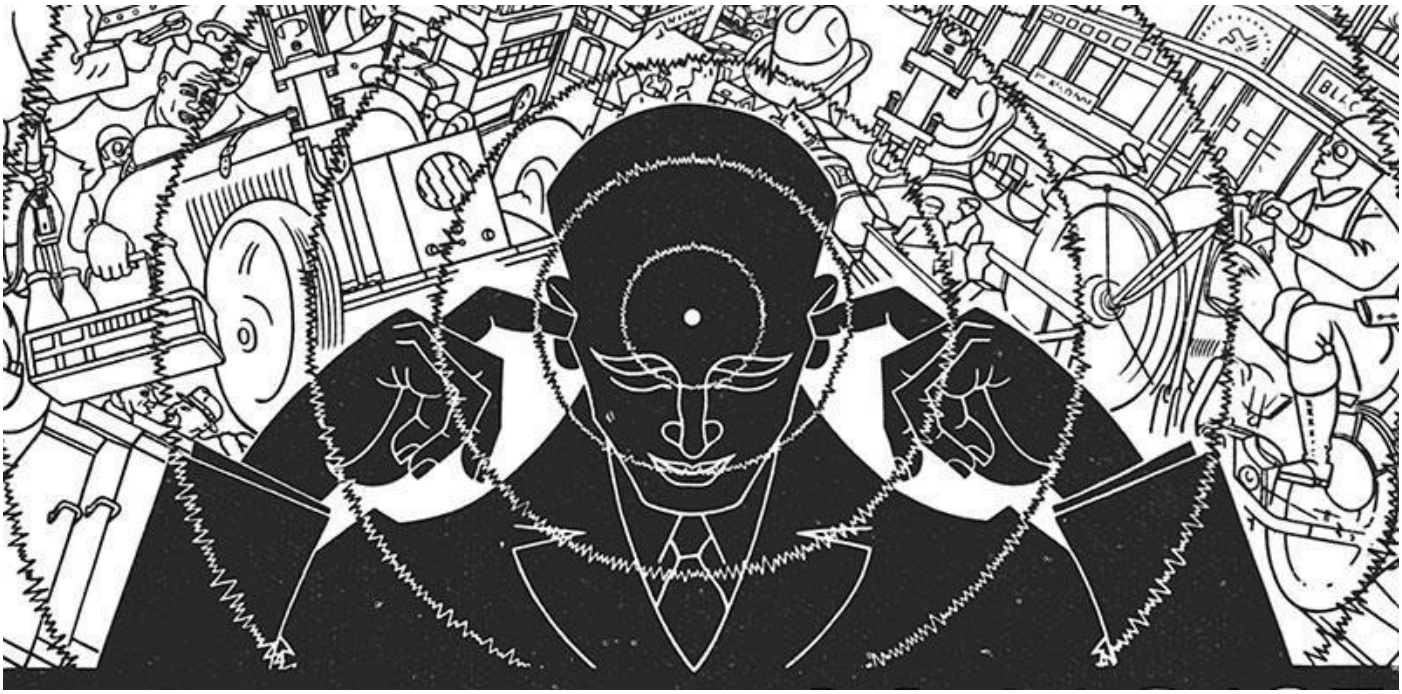
If a person is wearing headphones with a built-in microphone then any surrounding loudspeakers could be utilised to reinforce the output of the headphones, irrespective of whether they were listening to speech, music or sound effects. By altering the frequency response of the content in the loudspeakers so that it augments the headphone content by only transmitting any missing content. This would be easily achievable by delaying the headphone signals so that they are time aligned with the loudspeakers. If the headphone content was already sufficient then no external sources would be used at all. If an open ear design was used, or bone conduction where the pinnae were completely uncovered then the worn devices could be used to augment the content being transmitted. Poor acoustics or noisy environments could be improved by increasing the percentage and quality of the direct sound versus the indirect, without isolating listeners from the shared auditory environment.




When wearing headphones there are typically two major drawbacks, acoustic isolation and in head locatedness (IHL), where everything sounds like it is coming from inside your head. There is also a third much more dangerous issue, that of potential hearing damage. People tend to listen at higher sound pressure levels when wearing headphones than when using loudspeakers, even when they are aware of the potential hearing damage. Content can be designed to be split across loudspeakers, a gamer might experience all their avatar's sounds through headphones connected to their controller, with all of the other sounds coming through surround sound loudspeakers. General monitoring sounds, like confirming that something is midway through a task might emanate from the device performing the action, whereas an alert might be replayed through the closest device to ensure that the end user is aware of the issue and can make a decision whether to act or not.



Some auditory interference is essential for safety reasons. Alarms are the most obvious, fire and emergency service vehicles being the ones that probably affect the most people. A simple solution is to mute all other sound sources, just like an incoming telephone call mutes the music player in a smart phone. If this is not desirable to silence the audio content then the frequencies present in the alarm can be filtered out (equalised) so that the alarm cuts through more clearly, which will also make it easier to identify the spatial location. Less critical monitoring can also be integrated into a system such as when a washing machine has finished its cycle, or even a lack of sound such as when a child has fallen asleep.



These are only a few of the potential interactions that could help the world become a quieter place. The feedback loop of transmission and reception, with intelligent choices, made in real-time can minimise the impact of any sound source within a shared environment. DSP can be linked to almost any event, even health measurements so that if a source is having a negative impact then it can be adjusted until the affect is minimised. With wearable computing gradually becoming more popular, and the popularity of smart phones it will become unusual not to have a device listening out for your acoustic comfort.

© Iain McGregor All Rights Reserved Find more at [www.soundformovingpictures.com](http://www.soundformovingpictures.com)