Psychology Documentation

SOUND DESIGN

Contents [hide]

Why is sound important for virtual reality?

A basic expectation of virtual reality is to provide an immersive experience for the user. Along with visual elements, sound is a major factor contributing to how immersive an experience can be. A more immersive virtual reality experience will more effectively alter mood states.

Sound in virtual reality

A person experiencing a simulation is represented by an avatar within the virtual environment. They will experience visual information from the point of view of the camera and will experience sound information from an audio listener attached to the camera. The camera position and orientation are the location and direction from which the person's virtual ears will perceive the environment. With head-tracking, the simulation will know which direction within the virtual environment a person is facing. Tracking allows the distance and orientation to objects and sound sources within the virtual environment to be known. From this information, a sound simulation of what each ear should be hearing is generated. The user will detect spatial cues in the sound simulation, similar to how they are perceived in the real world.

Distance Cues

In this example the controllers are the ‘ears’ in the simulation.

About sound

When we hear sound, we perceive a disturbance caused by a vibrating object. The object can be tiny, like the buzzing wings of a mosquito, or huge, like a rumbling volcano. The vibrating object spreads waves of disturbance, called pressure waves, outwards through the medium around it. We are most familiar with sound travelling through air. When a pressure wave travels through air, there is a ripple of compressed air (high pressure) followed by less compressed air (low pressure). These are the peaks and troughs of the wave.

Sound waves are made up of compression and rarefaction (expansion).

When the wave reaches our ears, the strength, shape, frequency and duration of these ripples are clues we use to determine what we are hearing and where it is located.

Elements of sound

Frequency

How fast the wave is oscillating (alternating) between high and low pressure is referred to as frequency. The frequency of the oscillating high and low pressure is measured in hertz. One hertz is equal to one cycle of the oscillation in one second. The average adult human can hear sound waves with a frequency between 20 Hz and 16,000 Hz.

20Hz to 20kHz (Human Audio Spectrum)

Sound frequency example from 20 Hz to 20,000 Hz.

Pitch

Our ears perceive frequency over time as pitch. If a sound wave is high frequency, this will translate to a high pitch sound. Low frequency waves will be perceived as sounds that are lower in pitch. For example, a foghorn has a frequency of around 250 Hz, while the buzzing of a mosquito is around 400 Hz.

Sound Waves: High Pitch and Low Pitch

A demonstration of pitch.

High frequency is perceived as high pitch; low frequency is perceived as low pitch.

Amplitude

The power, or energy, carried by a sound wave is referred to as amplitude. Amplitude is measured by the difference in height of the peak and trough of the wave. As a sound wave travels through a medium, it will lose power and the wave height will reduce.

Low amplitude and low wave height; high amplitude and high wave height.

Timbre

A difference in material, shape and size of a vibrating object will send out differently shaped sound waves. Timbre is the tone quality that we hear from this difference in wave shapes.

Sound Made Simple - Sound & Hearing - Timbre

A short example of timbre

The change in shape of the sound wave over time is called the envelope. The envelope contributes to describing the timbre of a sound. For example, sound waves produced by a foghorn, shotgun and school bell will be different because they carry different vibrations of energy.

Examples of the differently shaped sound wave envelopes.

Timbre is important in distinguishing between competing sounds. It is a major factor contributing to why we can, for example, pick out the individual sounds of instruments playing in an orchestra when they seem come from the same distance and direction (Maddox & Shinn-Cunningham, 2011).

Different sounds have different waveforms.

Sound and spatial hearing

We gather separate sound information with our two ears first, then the brain processes that information. This is called binaural hearing and is important for determining sound within a three-dimensional environment (Vorländer & Sinn-Cunningham, 2014).

Binaural hearing requires sound information from the ears to be processed in the brain.

We use certain cues in the differences between the sound information received by each ear to determine the distance and direction of sounds, and to distinguish between sounds of similar distance and direction.

Binaural cues

Having two ears allows us to hear from two slightly different positions. When we compare what each ear perceives we pick up binaural cues. These cues help us build a picture of what we are hearing and where it is located. The slight difference in angle of each ear to the source of a sound will mean that a sound will arrive at one ear faster than the other. This is the main cue that allows us to locate the direction of a sound source (Vorländer & Sinn-Cunningham, 2014).

Sound Localisation | MED-EL

Demonstration of how binaural cues help with sound localisation.

A difference in angle will cause sound to arrive at each ear at different times.

Anechoic distance cues

When sound travels through a medium, the shape and strength of the sound wave will change. This allows us to pick up anechoic distance cues. The most notable change is a reduction in the sound intensity, or amplitude. Anechoic distance cues are most effective with sounds that we know. This is because we determine how much the characteristics of a sound changes as they travel through a medium. For example, if we hear a dog barking, we can determine if it is a dog barking in the distance, or a dog barking in the same room because of how much the sound has changed from what we recognise as the source. If we hear an unfamiliar sound, we can't determine the change in sound because we have nothing to compare it to. Yet even with unfamiliar sounds, anechoic distance cues can contribute to our understanding of the distance of a sound source (Mershon, 1997). This is because environmental factors, like turbulence, wind, humidity, temperature and atmospheric pressure, will change the characteristics of a sound wave as it travels to a distant listener.

Sound will change as it travels from its source.

(Vorländer & Sinn-Cunningham, 2014) Over distance, the shape of sound waves will change.

Reverberation

Another cue in determining distance is the reverberation of a sound source. Reverberation is the reflected energy from sound waves. When compared to the energy received directly from the source, reverberation can reveal how far away a sound source is located. When sound bounces off surfaces it creates a pattern of reverberation. This pattern of reverberation provides us with information that we can use to characterise of the listening environment. For example, we can determine how spacious a room is and from which materials it is made.

Reverberation is caused by sound reflecting off objects.

examples of reverb

Examples of reverberation in different spaces.

Almost always there is reverberation of a sound. An environment that lacks reverberation will seem unnatural.

Cross-modal perception

Other senses contribute to our understanding of sound. Visual and proprioceptive cues give us information about spatial location. These cues are used when processing auditory cues to build a sense of space. Integration of this information can help form better spatial awareness and is called cross-modal perception. When we detect cues from different senses together, we perceive them as a single multisensory event (Popescu et al., 2014). When we have prior knowledge gathered from another sense, we can form more accurate spatial perception. Prior knowledge helps provide meaning in the form of context. Context helps us decide what new information is relevant, or related to, to prior information. For example, if you have seen a dog across the street and after turning away you hear a dog bark, the prior visual knowledge of the location of the dog will help you to use the sound information more accurately.

Auditory scene analysis

In everyday life many different sound sources contribute to what our ears perceive. With all the competing sounds reaching our ears, we need a way to distinguish sound sources. Our auditory system uses a process called auditory scene analysis to organise sound information in a meaningful way (Carlyon, 2004). We organise what we hear by grouping sound sources. We do this to sounds that are similar in direction and distance, and to sounds of similar frequency, timbre and duration. Determining the location, direction and quality of competing sound sources helps build a picture of the sound environment. We create a scene made up of sound sources. We process this scene by segregating sound sources that make up the scene. When we group sounds together that we perceive as belonging to the same source, we also separate that source from all others. Then we stream the sound source by focusing our attention on that source over time.

The Components of Sound | MED-EL

Example of different groups of sounds that make one sound source.

Simulating space with sound

Entire virtual environments can be created using only sound. Different levels of sound sophistication will have different levels of impact on a virtual experience. To use sound to simulate space in virtual reality, the binaural experience we have in the real world needs to be recreated. To do this, sound information must be created for two separate channels. These two channels replicate the waveforms of sound that our eardrums would receive in the real world. We can then feed the sound signals from the two channels, through speakers or headphones, to our ears (Vorländer & Sinn-Cunningham, 2014). Headphones are more effective because sound waves from speakers will overlap. This will cause the spatial cues to be mixed and present reverberation that does not represent the virtual environment to the listener (Cudworth, 2014).

Diotic sound

When identical sound information is sent to each ear it is called diotic. Diotic sound is perceived as 2-D sound because it gives no spatial information to the listener. A listener will perceive the sound inside their head as if the sound source is located midway between their ears.

Dichotic sound

We say sound is dichotic when slightly different sound is received by each ear. The difference in information carried by the sound is only enough to give the impression that a sound is to the left or right of a listener. This is usually done by delaying the arrival of a sound to one ear and increasing the sound volume to the other ear, depending on the location of the sound source. Dichotic sound won't provide enough information to locate sound sources within a virtual environment. A listener will experience the sound as being located within their head, and perceive the sound source located along a line somewhere between one ear and the other.

Diotic sound provides no spatial information; dichotic sound can convey a location to the left or right of the listener.

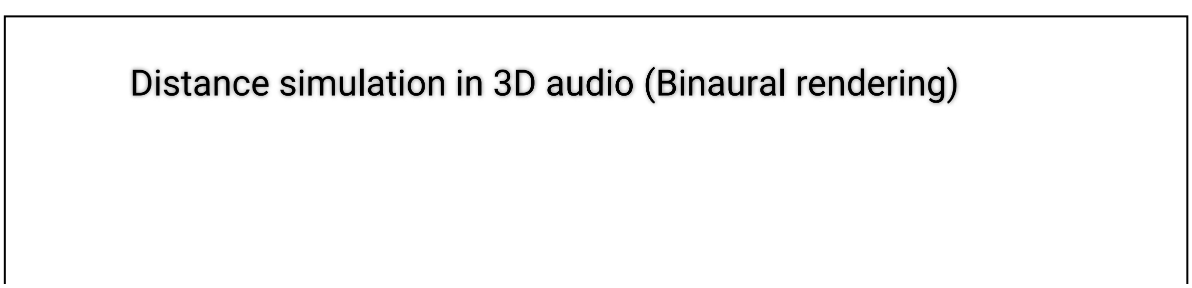
Spatialised sound

Truly 3-D sound, known as spatialised audio, takes advantage of binaural hearing. The location of sound sources within a virtual environment are used to generate the spatial cues each ear would receive in the real world. These generated cues will help a listener pinpoint the location of a sound source. Virtual reality gives the additional benefit of head-tracking. Head-tracking allows the user to move within the virtual environment and receive updated sound information depending on their orientation. When used with head-tracking, spatialised audio provides a very compelling representation of space (Langendijk & Bronkhorst, 2000).

How spatialised sound works in virtual reality.

Only spatialised audio will provide the necessary cues to determine the precise location of an object within virtual reality. Spatialised audio is required to give an object a spatial location relative to the person experiencing the reality.

Multiple sound sources from a virtual environment can be represented with spatialised sound.



An example of sound being moved with spatialised audio processing.

Virtual Barber Shop Hair Cut - 3D Sound (Amazing!!)

A virtual haircut experience using spatialised sound to create an auditory virtual environment.

Immersion and sound

Achieving an immersive experience is a priority when designing for virtual reality. To a large extent, being immersed in the experience will be determined by:

- how engaged a person feels with the virtual environment
- the fidelity of the sensory information from the environment.

Some key factors to consider when using sound in virtual reality are outlined below.

Ambient sound

Ambient sound is a part of our daily lives. It could be the hum of distant traffic in a city, the wind through trees in a forest, or the white noise of electrical appliances in a house. You may not notice when it is present, but you will notice when it is absent. It is sound that happens in the background. An absence of background noise will make it obvious that an environment is artificial. Background sound can enhance immersion within an experience. It provides a feeling of connection with the environment (Serafin & Serafin, 2004). If background noise is not included, it can reduce how immersive an experience is (Fencott, 1999).

Cross-modal enhancement

Sound can be used to indicate events that take place outside the field of view. If there are gaps in the spatial information from one sense, we can use the information from other senses to achieve a better understanding of space. We combine the sensory information to make a better spatial understanding of our environment (Ernst & Bülthoff, 2004). Our vision relies on a narrow field of light that reaches our eyes at any one time. Sound provides information from the environment that visual information can't provide. Sound comes to us from all directions at all times, even in complete darkness. We can rely on sound to provide us with information in situations where our vision is unable. Liminal's Ion experience uses cross-modal perception to enhance the emotional effect. An increase in the intensity of sound is accompanied by an increase in the complexity of visual elements.

Sound and interaction

When people perceive that they can affect or control their environment, they are said to have a sense of agency over the environment. Agency is the ability to initiate action and cause effect. This interaction with the environment will allow people to feel embodied within an experience. Sound that is the outcome of actions in a virtual world will reinforce the sense of control that the person has over the environment. If actions and outcomes are accompanied by sound, the combined effect enhances the sense of agency. A sense of agency will enhance immersion.

Sound and emotion

There are certain sound properties and sound listening conditions that have been studied and associated with emotion.

Binaural beats and emotion

A binaural beat is an illusion created by the brain. A tone of slightly different frequency played to each ear (via headphones) will cause the brain will perceive a new tone. The frequency of the new tone is equal to the frequency of the difference between the two tones. This new tone is a binaural beat.

Unisonic - Binaural Beats Demonstration

An example of binaural beats and music.

Binaural beats demonstration | 14 Hz and 6,5 Hz

Examples of binaural beats: 14Hz – related to awakeness and alert contentration on tasks. 6.5 Hz – related to deep meditation.

If a tone of 250 Hz is played through one headphone speaker, and a tone of 290 Hz is played through the other speaker, we will perceive a binaural beat as a tone of 40 Hz.

A binaural beat is perceived by the brain from two separate sounds.

Studies have shown the effect of binaural beats on emotional states. One experiment by Jirakittayakorn and Wongsawat (2017) investigated the effect of listening to tones that caused a binaural beat of 40 Hz for a period of 20 minutes. The participants experienced a significant decrease in tension, confusion and depression. They also experienced an increase in vigour. Another experiment exposed participants to the same binaural beat for a slightly longer period (30 minutes). These participants reported a significant increase in fatigue, yet no changes in other emotional states. The researchers concluded that a longer exposure caused an over-activation of the brain, resulting in fatigue. The shorter length of exposure had positive effects for an energised emotional state.

The table shows the results of 20 minutes of exposure to a 40 Hz binaural beat (Jirakittayakorn & Wongsawat, 2017).

Environmental sound and emotion

Sound and annoyance

Most psychological studies investigating the relationship between emotion and environmental sound have focused on noise annoyance (Berglund et al., 2002). Noise annoyance is an unpleasant feeling evoked by environmental sound. Evidence shows that this unpleasant feeling is a response to sound qualities found in stimuli perceived as loud, sharp and rough. Loudness is the subjective experience of sound intensity. People in the presence of high intensity sound will perceive the sound as loud. Sharpness is the sensation of a sharp, high-frequency sound. The right amount of sharpness added to a sound will give the sound a powerful characteristic. Too much sharpness will make a sound seem aggressive (Berglund et al., 2002). Roughness is related to sensory dissonance (Parncutt, 1989). It is the beating sensation produced by the interaction of components of physical sound in fast modulation. Tones of similar amplitude are more likely to produce a rough sound. Liminal's Ion experience uses a heavy mechanical sound that builds in volume as the central platform winds up to launch. Care was taken to avoid annoyance by using a sound that did not contain too much sharpness.

Nature sounds and emotion

Further evidence for the effect of environmental sound on emotion comes from studies of nature sounds such as wind, rain, waves and birds. In one study, faster task recovery was demonstrated by participants who had been exposed to nature sounds following a stressful task. In contrast, exposure to noise did not correlate with faster task recovery (Alvarsson et al., 2010). The positive effect of nature sounds does not require exposure to real-world nature settings. Annerstedt et al. (2013) showed that exposure to nature settings in virtual reality aids recovery from experimentally-induced stress. Crucially, recovery was faster for settings with nature sounds compared to settings without sound.

Colours of noise and emotion

Noise colour refers to differences in the power spectrum of noise signals. Different colours of noise will have different balances of power across the spectrum. The different balances are characterised as different colours and can be perceived by human ears, with some colours found to affect emotion.

The colours of noise

White noise is similar to television or radio static. It has constant and equal energy across the frequency spectrum, creating a flat sound. *Pink noise* has a frequency spectrum that decreases in intensity at a rate of three decibels per octave. This means that the higher frequencies of pink noise have less power than the lower frequencies. The sound of pink noise resembles a heavy and steady rain. *Brown noise* has a frequency spectrum with even less energy at higher frequencies than pink noise. When heard, it has a damped or soft quality. The sound is a low roar resembling an ocean or large waterfall.

Colors of Noise - Sound & Spectrum

Examples of different noise colours.

White noise and pain

Evidence suggests that white noise can be used to reduce pain and anxiety during minor surgical procedures. Gardner et al (1960) tested the effect of white noise on patients undergoing dental procedures that normally require sedation or anaesthesia. Results showed that white noise had calming and analgesic effects in 65 per cent of patients.

Noise colours, peripersonal space and emotion

Peripersonal space (PPS) refers to the emotionally-tinged zone around the human body that people experience as 'their space' and which others cannot intrude without arousing discomfort. Research has investigated whether approaching threatening sound stimuli influence PPS boundaries. Ferri et al. (2015) used emotion-inducing looming sound sources (noise colours) to investigate whether PPS representation was affected. Overall, brown and pink noises were judged as more arousing and less pleasant compared to white noise. Tajadura-Jiménez et al. (2010b) showed that unpleasant approaching sound sources evoke more intense emotional responses than receding ones. This approaching-receding difference was found, however, only for negative emotion-inducing sounds and not for neutral or positive sounds.

Self-representative sound and emotion

Self-representative sounds are sounds that we recognise as being made by our body. Self-representative sounds include footsteps, heartbeats and breathing. These sounds contribute to self-awareness and can produce emotional reactions in listeners. Tajadura-Jiménez and Västfjäll (2008) tested the effect of heartbeat sounds on participants' own heartbeat, following which they tested their emotional responses to pictures. Results showed a small but significant effect of heartbeat sounds on participants' own heart rate. The effect on emotional responses to pictures was much larger. Compared to baseline, pictures appeared more arousing following exposure to fast heartbeats, and less negative following exposure to slow heartbeats. Liminal's Ion experience uses drum beats to entrain heart rate. As new levels of the experience begin, the drumming sound increases in speed. A user's arousal will increase with the increase in heart rate. Liminal's Ripple Effect experience also uses sound to entrain heart rate. A heartbeat sound is reduced from 80 BPM to 60 BPM over the duration of the experience. A heartbeat sound will promote self-identification. The reduction in sound rate evokes a physiological response in users. A heart rate of 60 BPM is associated with a relaxing state. Development of Liminal's Ripple Effect revealed that some self-identifying sounds can be perceived as threatening. Early versions of Ripple Effect used realistic breathing sounds, yet these breathing sounds provoked feelings of anxiety in users. Subsequent versions used synthetic breathing sounds and were received positively by users.

Spatialized sound and emotion

Spatialized sound contributes to the realism of auditory virtual environments. It provides information about the distance of sound sources and the size and configuration of a listening environment. Early evidence for the emotional effect of spatialized sound comes from a study by Västfjäll et al. (2002). They found that preference for virtual sounds decreased with sound reverberation level. High reverberation was consistently judged more unpleasant than medium and low reverberation. These findings contrast with prior evidence from a study focusing on frequency and reverberation. The study found that reverberation at middle and high frequencies (500 Hz–1 kHz) made a concert hall feel lively, whereas short reverberation times (also at middle and high frequencies) made the hall feel deadened or dry. However, reverberation with frequency below 350 Hz made the hall feel warm. One possible explanation for Västfjäll et al.'s observation is that high reverberation produced a mismatch in perceived room size between vision and sound. If correct, this interpretation points to the importance of designing virtual environments where sensory cues are congruent across different modalities. In an experiment of simulated environments using reverberation, Tajadura-Jiménez et al. (2010a) observed a positive effect of low frequency virtual sound (262.6 Hz and 494.8 Hz) on emotion, arousal, and perceived pleasantness. The experiment involved simulating a concert hall, a semi-open inner courtyard, and a small studio. Two orientations were rendered in each room. One with the listener facing the source and one with the listener looking away from the source. Results showed that the concert hall and the courtyard were considered less safe and evoked more unpleasant and arousing responses than small spaces, which were perceived as calm, safe and pleasant. Moreover, sources located behind the listeners produced greater arousal than sources in front of the listeners. Liminal's Ion experience uses the sound of a torpedo that originates from behind the user and then projects forwards. The initial location of the sound is designed to heighten the user's arousal.

References

Alvarsson, J J, Wiens, S and Nilsson, M E, 2010. Stress recovery during exposure to nature sound and environmental noise, *International Journal of Environmental Research and Public Health*, 7(3):1036–1046.

Annerstedt, M, Jönsson, P, Wallergård, M, Johansson, G, Karlson, B, Grahn, P, Hansen, A M and Währborg, P, 2013. Inducing physiological stress recovery with sounds of nature in a virtual reality forest: results from a pilot study, *Psychology & Behavior*, 118:240–250.

Begault, D, Wenzel, E, Lee, A and Anderson, M, 2012. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, in *108th Audio Engineering Society (AES) Convention*, Vol. 5134.

Berglund, B, Hassmén, P and Preis, A, 2002. Annoyance and spectral contrast are cues for similarity and preference of sounds, *Journal of Sound and Vibration*, 250(1):53–64.

Carlyon, R P, 2004. How the brain separates sounds, *Trends in Cognitive Science*, 8:465–471.

Cudworth, A L, 2014. *Virtual World Design: creating immersive virtual environments* (Taylor & Frances: Boca Raton), pp 219–236.

Ernst, M O and Bülthoff, H H, 2004. Merging the senses into a robust percept, *Trends in Cognitive Sciences*, 8(4):162-169. Fencott, C, 1999. Towards a design methodology for virtual environments, in *Proceedings International Workshop on User Friendly Design of Virtual Environments UCDIVE'99*, University of York, England.

Ferri, F, Tajadura-Jiménez, A, Väljamäe, A, Vastrano, R and Costantini, M, 2015. Emotion-inducing approaching sounds shape the boundaries of multisensory peripersonal space, *Neuropsychologia*, 70:468–475. Gardner, W J, Licklider, J C R and Weisz, A Z, 1960. Suppression of Pain by Sound, *Science*, 132(3418):32–33.

Gardner, W. J., Licklider, J. C. R., & Weisz, A. Z. (1960). Suppression of pain by sound. *Science*, *132*(3418), 32-33.

Garner, T and Grimshaw, M, 2011. A climate of fear: Considerations for designing a virtual acoustic ecology of fear, in *Proceedings 6th Conference on Interaction with Sound*, pp 31–38. Grimshaw, 2009. The audio uncanny valley: sound, fear and the horror game, in *Proceedings Audio Mostly 4th Conference on Interaction with Sound*,

Glasgow. Hendrix, C and Barfield, W, 1996. The sense of presence in auditory virtual environments, *Presence*, 5(3):290–301. Jirakittayakorn, N and Wongsawat, Y, 2017. Brain responses to 40-Hz binaural beat and effects on emotion and memory, *International Journal of Psychophysiology*, 120:96–107.

Langendijk, E H and Bronkhorst, A W, 2000. Fidelity of three-dimensional-sound reproduction using a virtual auditory display, *Journal of the Acoustical Society of America*, 107(1):528–537.

Maddox, R and Shinn-Cunningham, B B, 2011. Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention, *Journal of the Association for Research in Otolaryngology*, 13:119–129.

Mershon, D H, 1997. Phenomenal geometry and the measurement of perceived auditory distance, in *Binaural and spatial hearing in real and virtual environments* (eds: R Gilkey and T

Anderson), Lawrence Erlbaum Associates: New York, pp 251–214.

Parncutt, R, 1989. *Harmony: A Psychoacoustical Approach*, Berlin: Springer-Verlag. Popescu, G V, Trefftz, H and Burdea, G C, 2014. Multimodal interaction modelling, in *Handbook of Virtual Enviro*nments (eds: K S Hale and K M Stanney), CRC Press: Boca Raton, FL, pp 411–434.

Serafin, S and Serafin, G, 2004. Sound design to enhance presence in photorealistic virtual reality, in *Proceedings of the 2004 International Conference on Auditory Display*, Sydney (eds: S Barrass and P Vickers), pp 1–4).

Slater, M and Usoh, M, 1994. Body centred interaction in immersive virtual environments, in *Artificial Life and Virtual Reality* (eds: N Magnenat Thalmann and D Thalmann), John Wiley and Sons, pp 125–148.

Storms, R L and Zyda, M J, 2000. Interactions in perceived quality of auditory-visual displays, *Presence: Teleoperators & Virtual Environments*, 9(6):557–580.

Tajadura-Jiménez, A, Larsson, P, Väljamäe, A, Västfjäll, D and Kleiner, M, 2010a. When room size matters: Acoustic influences on emotional responses to sounds, *Emotion*, 10(3):416–422.

Tajadura-Jiménez, A, Väljamäe, A, Asutay, E and Västfjäll, D, 2010b. Embodied auditory perception: The emotional impact of approaching and receding sound source, *Emotion*, 10(2):216–229.

Tajadura-Jiménez, A and Västfjäll, D, 2008. Auditory-induced emotion: A neglected channel for communication in human-computer interaction, in *Affect and Emotion in Human-Computer Interaction*, pp 63–74.

Västfjäll, D, Larsson, P and Kleiner, M, 2002. Emotion and auditory virtual environments: affect-based judgments of music reproduced with virtual reverberation times, *CyberPsychology & Behavior*, 5(1):19–32.

Vorländer, M and Sinn-Cunningham, B, 2014. Virtual auditory displays, in *Handbook of Virtual Environments: Design, Implementation, and Applications* (eds: K S Hale & K M Stanney), CRC Press, pp 87–114.