Distance in audio for VR: constraints and opportunities

A. McArthur
Centre for Digital Music, Media & Arts Technology, School of Electronic Engineering & Computer Science
a.mcarthur@qmul.ac.uk

M. Sandler
Centre for Digital Music, School of Electronic Engineering & Computer Science
mark.sandler@qmul.ac.uk

R. Stewart
Centre for Digital Music, School of Electronic Engineering & Computer Science
rebecca.stewart@qmul.ac.uk

Figure 1: Overhead shot of the moving set for the cinematic VR film 'Sundowning'

ABSTRACT

Spatial audio is enjoying a surge in attention in both scene and object based paradigms, due to the trend for, and accessibility of, immersive experience. This has been enabled through convergence in computing enhancements, component size reduction, and associated price reductions. For the first time, applications such as virtual reality (VR) are technologies for the consumer. Audio for VR is captured to provide a counterpart to the video or animated image, and can be rendered to combine elements of physical and psychoacoustic modelling, as well as artistic design. Given that distance is an inherent property of spatial audio, that it can augment sound's efficacy in cueing user attention (a problem which practitioners are seeking to solve), and that conventional film sound practices have intentionally exploited its use, the absence of research on its implementation and effects in immersive environments is notable. This paper sets out the case for its importance, from a perspective of research and practice. It focuses on cinematic VR, whose challenges for spatialized audio are clear, and at times stretches beyond the restrictions specific to distance in audio for VR, into more general audio constraints.

AM '17, August 23–26, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5373-1/17/08.

KEYWORDS

Immersive audio, cinematic virtual reality, auditory distance, audiovisual perception, spatial sound rendering.

CCS Concepts • General and reference → Document types → Surveys and overviews

ACM Reference format:
A. McArthur, M. Sandler and R. Stewart. 2017. Distance in audio: constraints and opportunities. In *Proceedings of AM '17, London, United Kingdom, August 23–26, 2017*, 7 pages.
https://doi.org/10.1145/3123514.3123530

1 INTRODUCTION

Although not a new medium, VR is perhaps the most obvious of the immersive trends [1]. The level of impact that audio in VR can achieve with relatively low computational cost is promising, and continues to improve. Google Cardboard's spatial audio for example, computes audio in real-time yet is implemented on mobile devices whilst "..*most of the processing takes place outside of the primary CPU*"[2]. Their distance encoding includes amplitude differences, 'smooth-changing' low pass filtering, early reflections, occlusion and a dedicated reverberation engine, which takes account of environmental properties, and creates reverberation that decays differently for different frequencies [3].

Dynamic binaural synthesis, utilizing head-related transfer functions (HRTFs), virtual loudspeakers and headset orientation data, provides a compelling experience (Fig. 2) which can be delivered (minimally) over standard headphones, a smartphone and cardboard headset. This affords a sounding world where distance, location and environmental cues remain independently static (or dynamic) when we move and it serves to reinforce presence [4,5] which in turn can 'uplift' potentially presence-breaking features of image [6,7].

Figure 2: VR user playback with professional headphones and an Oculus DK2.

Yet the ease of consumption does not reflect the challenges faced by content creators. With sound design having inherited its (relative to image) low status from existing film and game conventions, it currently remains a largely post-production concern. There are a further and distinct range of constraints practitioners face (directly or indirectly) when attempting to add distance.

This paper draws from academic research as well as the first author's recent cinematic VR film 'Sundowning' (see Fig. 1) which is currently in post-production. This short form film, directed by Angela McArthur, Shivani Hassard and Vanessa Pope, is being made by practitioner-researchers in partnership with industry. It was, and continues to be an ambitious 'one-take' project involving a moving set, interactive sound design, and computer generated images.

Sound design for cinematic VR differs from conventional filmmaking most fundamentally in its requirements for spatial diegetic sound (ideally in three dimensions), and the hiding of microphones (with 360^{o} cameras there is no 'frame' outside of which to place boom or other microphones). These requirements can be attended to during capture, and post-production. The net effect of these differences however, is an increased demand for time and expertise.

At inception, 'Sundowning' gained an award for its sound design treatment. It depicts Alzheimer's from a first-person perspective, aiming at an imaginative, insightful portrayal of the condition. It was produced under very challenging temporal and environmental constraints, and these constraints impacted the realisation of the project, including planned sound design. One of the four actors involved was a five-year-old child who did not attend rehearsals. The adult actors had a limited rehearsal period, and were unfamiliar with the format (with directives to treat the camera as a character, with the significance of scale, and with stitch lines). Anxieties arose from the peculiarities of the performing the piece in one take.

Pre-production on the film involved material and simulated models of the set, experimentation with actors using different sound-producing objects (props, costume), the creation of a click track system (abandoned during production due to time constraints) which could synchronise the actors over multiple takes, thus creating an atypical opportunity for cutting *between* takes, whilst also affording the design of a rhythmic 'signature'. The decision to use standard sound capture and rendering techniques meanwhile, would create contrast with the interactive audio. This was designed to be triggered if the user looked directly at a particular character, during a particular monologue.

These conditions are not detailed to set this production up as unusual - productions generally operate under such sub-optimal conditions. It is important to qualify the realities of a production environment at the outset. Much research relevant to distance cues in immersive audio is, quite reasonably, carried out under controlled experimental conditions. There are stark differences in the aims and outcomes of these approaches, and in the forms of theoretical and practical knowledge produced as a result. It is the position of this paper (as well as its hope) that both practice and research will be instrumental in understanding and applying distance to audio for VR. This poorly understood, implemented and utilized cue is apt to be given attention from all perspectives.

2 THE CONSTRAINTS & OPPORTUNITIES

2.1 Engineering

2.1.1 Audio capture is a challenge for any production. In conventional filmmaking, automated dialogue replacement (ADR), where actors overdub their dialogue in a recording studio separately to filming, is common practice and proves effective against on-set recording in many cases, and for a variety of reasons. It becomes more complicated and thus less economic for spatial audio practice, where room reflectance, spatial impression, and other cues need to be matched to a scene, and where a lack of location availability may create the need for (often manual) synthesis. This risks audio element mismatching, unless all audio is similarly processed or other workarounds are found, which is time-consuming:

> *"We went to great efforts to match the sound of the ADR with the location dialogue using cabin impulse responses captured on the plane, and we put every voice through an individual HRTF panner to match the exact position on screen. It also proved quite challenging to match the sound of the last scene, shot in a studio, with those shot in the actual cabin because of the higher ceiling in the studio".* [8: 4]

Such practices have a more compound impact on delivery schedule and cost of production than has been the case for conventional filmmaking. That said, sound is more amenable to post-production than image, so performances are chosen for their visual success, often leaving sound with some ground to recover after capture. For 'Sundowning' it was hoped that ADR could be avoided, however due to ambient noises (low intermittent rumbles of passing trains, hum from overhead lighting, high volume broadband sounds due to the moving set) and the choice for visual success in performances, ADR was required.

2.1.2 Audio rendering developments are in an exciting, formative phase, with many spatial sound rendering solutions available. Plugins for digital audio workstations (DAWs) as well as games engines and audio middleware, are increasingly being made available free of charge to encourage content production, a recent opportunity. These provide varying degrees of fidelity, interoperability and capability in the organisation, placement and transformation of sound sources.

For scene-based (ambisonic) recordings commonly used in cinematic VR production, which represent the spatial wave front with spherical harmonics decomposition, certain transformations can be achieved (such as rotation) making them an appropriate choice for cinematic VR. However, object-independent distance encoding cannot be achieved. A 'zoom' or 'dominance' effect is possible on a soundfield representation, which does not discriminate individual sound events or separate sources. Effectively this will increase the gain level for elements of the soundfield in the direction or area of interest, and possibly also decrease the gain level for elements outside of it. Harpex claims to be able to isolate sound sources from a sound field by 'sharp spatial filtering'. Their 'post production shotgun' (their name for the Harpex acoustic zoom effect):

> *"...mimics the kind of spatial analysis that a human listener uses to separate and localize the many sound sources in a complex sound scene. This results in a level of channel separation and spatial definition that has never been possible before."* [9]

Work has also been carried out to alter the signal encoding and decoding processing, to achieve distance encoding, as well as to compensate for certain near-field effects [10 – 12]. Research continues to help shape the suitability of soundfield recordings for immersive audio, working towards source-separation specific to the ambisonic format [13]. Such continued research attests to the fact that a scene-based approach is less complex to render (relative to a object based approach, see below), and can offer a high fidelity of spatial impression for higher order formats. For such reasons, it is likely to remain a popular choice within a practitioner's toolset.

A blend of close mic mono or stereo recordings are often combined with a sound field representation and convolved with a variation of impulse responses (IRs) or HRTFs for binaural rendering. The different source recordings can be treated somewhat separately. However, to manually arrange, mix, and spatialize these sources, and then adjust them to dynamically synchronise with the moving image across the length of the piece is laborious, and unfeasible for many productions. Those practitioners from non-game audio backgrounds are often already upskilling to meet the technical demands required for spatial audio production and outputting, demands which are in a state of flux due to the rate of developments in this nascent ecosystem.

2.1.3 Object based audio systems are worth a mention. Whilst object-based systems are generally over-specified for cinematic VR, content with limited interactivity can employ an object-based approach to create offline mixes, to export [14]. Thus, their use for cinematic VR is feasible as a rendering, if not playback, tool. Interactivity can be added to such a mixed file though the use of a games engine, which treats sounds as objects, to have headset orientation data (certain polar coordinates in the spherical video) trigger sound events when looked at by a user, as used for 'Sundowning'. These coordinates can be manually mapped to an event in cinematic VR, and in this way it can be somewhat interactive, less distinct from real-time rendered interactive experiences. Object based systems effectively postpone the mix until exhibition (sound files and sound properties are held as objects and later assembled in three-dimensional space from their metadata) and some of these features are available to linear VR experiences, though a fuller dynamic modification of sound objects is not possible. Object based systems do require additional attention - each sound source in a scene must be recorded or synthesised, specified with properties and animated over time. Metadata needs to be generated (this may be possible automatically in some cases but can itself be a challenge [15]). Yet the return on investment for certain content is evident.

2.2 Perception

2.2.1 Distance perception is one of the least understood auditory perception processes, which might explain in part the lack of consensus on experimental design in its research. If we compare auditory distance to auditory direction, we can observe a clear imbalance in the consideration given to this aspect of sound localization.

There are a range of cues which we employ to judge auditory distance, the reliability of which vary with the stimulus, with sound source direction, and with sound properties of the environment (for a review see [16, 17, 18]). Our judgements are impacted by visual information, by experimental design (for example the measure of absolute or relative distances, the order of presentation), and non-perceptual factors (for example, our level of overall arousal, or our assessment of valence for a sound event [19]). They tend to be systematically biased inaccurate (most accurate at distances of approximately 1m), and imprecise, which is not the case for visual distance judgements [18, 20].

Our understanding further diminishes when examining an audio-visual environment, and again when that environment is rich and virtual, as is the case for VR. Work [21] has shown for example, that headsets induce under-estimation of distance, which already shows signs of 'compression' (i.e. we underestimate sound source distances in the far field, effectively bringing them toward us, and overestimate them in the near field, effectively pushing them away from us as if with a compressive power function [18]).

Auditory cues include:

- Intensity
- Temporal delay
- Frequency spectrum at near and far distances
- Inter-aural level differences
- Inter-aural time differences
- Direct-to-reverberant energy ratio (early reflections in particular)

2.2.2 Distance cues may be robust to spatial rendering and playback configuration. One study [22] has demonstrated that head movements and non-individualized HRTFs do not play an important role in the auditory perception of distance over loudspeakers or (static or dynamic) binaural (headphone) presentation. It should be noted however, that this work was not extended to audio-visual *or* virtual environments. Gerzon [23] advocated for artificial distance cues (such as early reflections) for this reason, combined with more control, less computation and undesirable natural artefacts, all afforded by synthesis.

2.2.3 Auditory capture. Although visual 'capture' can lead to the biasing of auditory distance judgements [24 – 26] visually degraded stimuli such as that in cinematic VR, may benefit from the uplifting effects of well-produced spatial audio producing a kind of auditory capture [27].

2.2.4 Familiarity of sources is an important consideration, with absolute judgements of distance relying on these. Listeners compare the spectral content and sound intensity of a sound event with their internal estimates of the source. Yet even in the case of the human voice, discrepancies might arise from the combination of unusual *types* of speech at their usual levels [28]. The extent of familiarity may be an important factor, and may explain certain exposure effects (see below). There exists an opportunity for VR to build on film sound practices which privilege speech intelligibility over spatial authenticity, as well as for creative exploration of this phenomenon. Familiar sound sources could be transformed in post-production to be less familiar (for example, physically less familiar via synthesis or semantically less familiar by – for example – creating a mismatch in speech qualities of a shouted event which is rendered or recorded with the qualities of a whispered event (less energy in the high frequencies, less overall intensity, etc). Inversely, unfamiliar sources could be convolved with familiar source properties. These exploratory ideas open up a world of creative possibilities for sound design in general, and spatial orientation effects specifically.

2.2.5 Exposure and training effects for distance perception are important though largely unchartered areas in research for VR. Work has shown that repeated exposure to unfamiliar sounds increases the accuracy of auditory distance estimates [29, 30]. Longitudinal work seems highly appropriate for dealing with VR as an emerging technology, and auditory distance as an under-represented field of enquiry.

2.2.6 Scientific paradigms that deal with auditory perception are changing in ways which may be helpful to understanding auditory distance perception, as a form of discrimination which employs a range of complex information in varying ways. In real environments, we combine multiple auditory (and often other sensory) cues to assess distance. Zahorik [31] demonstrated perpetual cue weighting was performed, with the primary cues being intensity and direct-to-reverberant energy ratio, though this varied with stimulus type, source direction, distance, background noise and other source or environmental properties. This approach takes into account prior information available to listener about the sound event. It builds upon research which demonstrates how we perceptually favour reliable sensory information to resolve cross-modal conflict [32], by suggesting that consistency is positively weighted, and

> *"...can perhaps explain how auditory distance perception processes are able to produce relatively stable– albeit biased - estimates of source distance under a wide range of acoustic conditions in which different distance cues are available to listeners."* [18: 417]

Such probabilistic perspectives have gained popularity in recent years as notions of sensory dominance fade. The Maximum Likelihood Estimation (MLE) model which assumes that the perceptual weighing process is statistically optimal, is proving particularly popular [33].

2.3 Design

2.3.1 Microphone placement is critical to the end user experience of cinematic VR content. For 'Sundowning' many periods across several days were spent on positioning the camera as evidenced by Fig. 3, 4 and 5, which show the progressive refinement of this process (with camera no rig; with camera and rig; with camera, rig and actor). For microphone placement, testing, and peak monitoring, less than 1.5 hours was taken. Actors did not rehearse with the chosen mics:

Figure 3: Camera position calibration for 'Sundowning' with camera no rig.

- 1 x Soundfield ST450 - positioned as close to camera as possible whilst as invisible as possible
- 4 x DPA 4060 capsule radio microphones - fitted as invisibly as possible to actors

It would be unfair to say that microphone placement requires as much consideration as camera, but the difference is stark.

Figure 4: Camera position calibration for 'Sundowning' with camera and rig.

Opportunities for the design of microphone placement seem to involve creative solutions to hiding or incorporating microphones on location, in a set, or on actors, which have little affirmative impact on the design outcomes. They are essentially risk-minimization measures, though the continual improvisation required is enlivening, and can renew one's enthusiasm for live sound recording. For manufacturers, microphone design is the design opportunity. Some manufacturers are focusing on all-in-one audiovisual solutions [34, 35] to lessen practitioner headaches, some are targeting price point and / or portability [36, 37].

2.3.2 Creating empathy and intimacy for characters. The closeness (to camera for VR) associated with these effects is often not achievable in production. To avoid distortion, stitch-line irregularities and adverse scaling effects, actors cannot move too close to the camera. Given close-mic recording practices, designing sound for these objectives is helpful, and may be less work than applying more cues to achieve greater distance.

Figure 5: Camera position calibration for 'Sundowning' with camera, rig and actor.

2.3.4 Creating a sense of depth is an obstacle for 'flattened' spherical video. This is aggravated by the quality degradation which video undergoes during cinematic VR rendering and playback. Research has pointed towards the fact that an "*overall sense of* [auditory]*distance is mainly dependent on the focal length* [of the camera]" [38: 1].

2.3.5 Attentional cueing is one of the most apparent opportunities for the use of auditory distance cues in VR, as a medium which affords users freedom of head movement (and thus attention). This has been a conundrum for practitioners, uneasy about relinquishing narrative control. Auditory perception research establishes sound's effectiveness to cue user attention consistently, even for narratively impoverished (experimentally controlled) stimuli ([6] Fig. 6).

3 CONCLUSIONS

Our privileging of consistency in perceptual representations from different senses which conflict (cross modal conflict) is actually useful for the practice of VR, enabling a coherent perceptual experience which tolerates disparity. We may discover this through trial-and-error, but VR production is costly, cumbersome, and intolerant to experimentation. At a time when conventions for capture and render remain in flux, and VR industries are yet to see solid returns on sizeable investments, research-informed practice is vital. Research too can learn from the use of ecologically valid stimuli and conditions, whose tacit knowledge may not always be articulated. Whilst the challenges for capturing, rendering, and perceiving distance in audio for VR abound, there are opportunities for researchers and practitioners to tackle the lack of available applied research in this area, and a certain freedom from convention in doing so.

Distance is an integral cue for our processing of sound in real and virtual environments, so this may be a necessity as well as an opportunity. The notion of audio depth of field, which the camera has yet to attain in VR, is an exciting one, and may help relocate audio's salience relative to image. VR may contribute to this by its popularity, its relative privileging of audio, and its ability to straddle the real and mediated immersive experience. Its potency in this respect, taken with the high degree of control that can be exerted in the virtual environment, make it a useful and ecologically valid tool for experimentation. Paradoxically perhaps, it may help us relinquish the controlled experimental paradigm in favour of a messy, rich combinatorial framework.

Figure 6: User view of research stimuli (5 identical bell alarm clocks).

ACKNOWLEDGMENTS

The authors would like to acknowledge support from the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology through Queen Mary, University of London.

REFERENCES

[1] Gartner. Hype Cycle for Emerging Technologies. 2016. Available at https://www.gartner.com/doc/3383817 [Accessed 06.2.17]

[2] Martz. N. Spatial audio comes to the Cardboard SDK. 2016. Available at https://developers.googleblog.com/2016/01/spatial-audio-comes-to-cardboard-sdk.html [accessed 22.03.17]

[3] Kelly, Ian. Spatial Audio and Immersion: VR's Second Sense. Presentation at Google I/O conference. 2016

[4] Larsson P. Västfjäll D. and Kliener M. Spatial auditory cues and presence in virtual environments. 2005. Submitted for *International Journal of Human-Computer Studies*. 2005

[5] Begault D. 3-D sound for virtual reality and multimedia. San Diego: Academic Press Professional, Inc 1994

[6] Clark. A. Cross-modal cuing and selective attention. 2005. In *Proceedings of Individuating the Senses* Glasgow, Scotland: University of Glasgow 2005.

[7] McArthur A. Disparity in Horizontal Correspondence of Sound and Source Positioning: The Impact on Spatial Presence for Cinematic VR. In Audio Engineering Society Conference: 2016 *AES International Conference on Audio for Virtual and Augmented Reality* 2016 Sep 21. Audio Engineering Society.

[8] Björling O, Thorsell E. Crafting Cinematic High End VR Audio for Etihad Airways. In Audio Engineering Society Conference: 2016 *AES International Conference on Audio for Virtual and Augmented Reality* 2016 Sep 21. Audio Engineering Society.

[9] Harpex. Statement in Gearslutz online forum. 31.03.11. Available at https://www.gearslutz.com/board/6490065-post1.html [Accessed 15.05.17]

[10] Daniel J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In Audio Engineering Society Conference: 23rd International Conference: *Signal Processing in Audio Recording and Reproduction* 2003 May 1. Audio Engineering Society

[11] Penha R. Distance encoding in ambisonics using three angular coordinates. In SMC: *Sound and Music Computing Conference 2008* July 31

[12] Adriaensen F. Near field filters for higher order Ambisonics. 2006.

[13] Perez, Andres. Current PhD student at Universitat Pompeu Fabra. Barcelona. Spain. Supervised by Emilia Gómez

[14] Pike C, Taylor R, Parnell T, Melchior F. Object-Based 3D Audio Production for Virtual Reality Using the Audio Definition Model. In Audio Engineering Society Conference: 2016 *AES International Conference on Audio for Virtual and Augmented Reality* 2016 Sep 21. Audio Engineering Society.

[15] J. Riedmiller, N. Tsingos, and S. Mehta, P. Boon. Immersive & Personalized Audio: A Practical System for Enabling Interchange, Distribution & Delivery of Next Generation Audio Experiences. SMPTE 2014 Annual Technical Conference & Exhibition

[16] Coleman PD. An analysis of cues to auditory depth perception in free space. Psychological Bulletin. 1963 May; 60(3):302.

[17] Kolarik AJ, Moore BC, Zahorik P, Cirstea S, Pardhan S. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Attention, Perception, & Psychophysics. 2016 Feb 1;78(2):373-95

[18] Zahorik P, Brungart DS, Bronkhorst AW. Auditory distance perception in humans: A summary of past and present research. Acta Acustica united with Acustica. 2005 May 1;91(3):409-20.

[19] Gagnon KT, Geuss MN, Stefanucci JK. Fear influences perceived reaching to targets in audition, but not vision. Evolution and Human Behavior. 2013 Jan 31;34(1):49-54.

[20] Da Silva JA. Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods. The American Journal of Psychology. 1985 Apr 1:119-44.

[21] Creem-Regehr SH, Willemsen P, Gooch AA, Thompson WB. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. Perception. 2005 Feb;34(2):191-204.

[22] Kearney G, Liu X, Manns A, Gorzel M. Auditory Distance Perception with Static and Dynamic Binaural Rendering. In Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet 2015 Mar 6. Audio Engineering Society.

[23] Gerzon MA. The design of distance panpots. In Audio Engineering Society Convention 92 1992 Mar 1. Audio Engineering Society.

[24] Gardner MB. Proximity image effect in sound localization. The Journal of the Acoustical Society of America. 1968 Jan;43(1):163-.

[25] Mershon DH, Desaulniers DH, Amerson TL, Kiefer SA. Visual capture in auditory distance perception: Proximity image effect reconsidered. Journal of Auditory Research. 1980 Apr.

[26] Zahorik P. Estimating sound source distance with and without vision. Optometry & Vision Science. 2001 May 1;78(5):270-5.

[27] Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. Current biology. 2004 Feb 3;14(3):257-62.

[28] Blauert J. Spatial hearing: the psychophysics of human sound localization. MIT press; 1974

[29] Coleman PD. Failure to localize the source distance of an unfamiliar sound. The Journal of the Acoustical Society of America. 1962 Mar;34(3):345-6.

[30] Mershon DH, Ballenger WL, Little AD, McMurtry PL, Buchanan JL. Effects of room reflectance and background noise on perceived auditory distance. Perception. 1989 Jun;18(3):403-16.

[31] Zahorik P. Assessing auditory distance perception using virtual acoustics. The Journal of the Acoustical Society of America. 2002 Apr;111(4):1832-46

[32] Spence, C & Driver, J. Attracting attention to the illusory location of a sound: reflexive cross-modal orienting and ventriloquism. 2000 in Neuroreport vol. 11 pp. 2057–61

[33] Mendonça C, Mandelli P, Pulkki V. Modelling the Perception of Audiovisual Distance: Bayesian Causal Inference and Other Models. PloS one. 2016 Dec 13;11(12):e0165391.

[34] Nokia Ozo https://ozo.nokia.com/eu/

[35] Visisonics RealSpace™ Audio Panoramic Camera http://visisonics.com/visisonics-audiovisual-camera/

[36] Core Sound Tetramic http://www.core-sound.com/TetraMic/1.php

[37] Senheiser Ambeo http://en-uk.sennheiser.com/microphone-3d-audio-ambeo-vr-mic

[38] Kruszielski LF, Kamekawa T, Marui A. The Influence of Camera Focal Length in the Direct-To-Reverb Ratio Suitability and Its Effect in the Perception of Distance for a Motion Picture. In Audio Engineering Society Convention 131 2011 Oct 19. Audio Engineering Society.