An open-source audio renderer for 3D audio with hearing loss and hearing aid simulations

Maria Cuevas-Rodriguez, Daniel Gonzalez-Toledo, Ernesto Rubia-Cuestas, Carlos Garre, Luis Molina-Tanco, Arcadio Reyes-Lecuona, David Poirier-Quinot, Lorenzo Picinali

Audio Engineering Society

Convention Paper

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

An open-source audio renderer for 3D audio with hearing loss and hearing aid simulations

Cuevas-Rodriguez, Maria[1]; Gonzalez-Toledo, Daniel[1]; de La Rubia-Cuestas, Ernesto[1]; Garre, Carlos[1]; Molina-Tanco, Luis[1]; Reyes-Lecuona, Arcadio[1]; Poirier-Quinot, David[2], and Picinali, Lorenzo[2]

[1] *University of Malaga, Avda. Cervantes, 2. 29071 Malaga (SPAIN)*

[2] *Imperial College London, Dyson School of Design Engineering, 10 Princes Gardens, SW7 1NA London (UK)*

Correspondence should be addressed to Cuevas-Rodriguez, Maria (mariacuevas@uma.es)

ABSTRACT

The EU-funded 3D Tune-In (http://www.3d-tune-in.eu/) project introduces an innovative approach using 3D sound, visuals and gamification techniques to support people using hearing aid devices. In order to achieve a high level of realism and immersiveness within the 3D audio simulations, and to allow for the emulation (within the virtual environment) of hearing aid devices and of different typologies of hearing loss, a custom open-source C++ library (the 3D Tune-In Toolkit) has been developed. The 3DTI Toolkit integrates several novel functionalities for speaker and headphone-based sound spatialisation, together with generalised hearing aid and hearing loss simulators. A first version of the 3DTI Toolkit will be released with a non-commercial open-source license in Spring 2017.

1 Introduction

This paper reports on an open-source C++ library developed within the EU-funded project 3D Tune-In [1] [2]: the 3DTI Toolkit. We report on the rationale behind developing a custom library, which was preferred to using existing solutions. We compare the alternatives and present the details of the real-time binaural and loudspeaker spatialisation, the hearing loss simulation and the hearing aid simulation implemented in the Toolkit.

The 3D Tune-In project (http://www.3d-tune-in.eu/) introduces an innovative approach using 3D sound, visuals and gamification techniques to support people using hearing aid devices. The consortium in the project has joined forces to develop the 3DTI Toolkit, a custom, open-source, multi-platform C++ library to respond to a challenging set of requirements regarding real-time performance and portability. Several applications are being developed using the Toolkit, deployed on multiple platforms (e.g. mobile, desktop or browser), tailored to different target audiences (e.g. older users or children) and scenarios (e.g. music listening or noisy environment simulation). The principal use cases are intended to train users in understanding and using the various settings of their hearing aids to attain optimum performance in different social contexts. The Toolkit provides a high level of realism and immersiveness within 3D audio simulations (both speaker and headphones-based), while allowing for the emulation (within the virtual environment) of hearing aid devices and of different typologies of hearing loss.

The hearing loss and hearing aid simulations – which were core requirements of the 3D Tune-In project – are novel and, to the best of our knowledge, not included in the available open-source 3D audio rendering systems.

Some of the hearing aid simulation features are strongly coupled with the spatialisation process. This was one of the main motivations for developing a custom library instead of using existing spatialisation solutions. The other motivation was the need of many features which, all together, were not found to be available in other existing libraries:

- Support for multiple platforms, including mobile and web.
- Full 3D positioning and movement of sources and listener.
- Customization of HRTFs.
- Spatialised reverb simulation.
- Support for both binaural and loudspeakers-based spatialisation.
- Open source.

The following sections present a brief technical background and a discussion about related work. Section 4 gives details on the supported features of the Toolkit, while Section 5 presents a measure of the performance of the Toolkit for soft real-time applications. Finally, Section 6 concludes the paper by laying out the plans for public release and licensing of the Toolkit.

2 Technical background

Binaural technology is not particularly recent [3]. The first binaural recording goes back to the end of the 19th Century, the *Théâtrophone* [4]. However, it is only within the last twenty years that the increase in the calculation power of personal computers enabled a complete real-time simulation of three-dimensional sound-field over headphones [5]. The aim of the binaural audio simulation is to provide the listener (through standard headphones) with the impression that sound is positioned at a specific point in three-dimensional space. The 3D characteristics of the sound can be captured during recording with special hardware, or simulated in post-production via spatialisation techniques. One of such techniques is binaural spatialisation. Generating binaural signals is based on convolving a monaural signal with head-related transfer functions (HRTFs), which model the directional filtering of the incoming signal due to the properties of the listener's body shape (i.e. torso, head and pinna) [6]. To select an HRTF for a specific listener, most 3D audio spatialisers obtain the HRTF from a public database, such as IRCAM Listen [7], MIT KEMAR, [8] or CIPIC HRTF Database [9]. The Toolkit has been tested with a subset of the IRCAM Listen database optimized by Katz [10]. In order to further individualize the HRTF to the listener, a customization of the Interaural Time Difference (ITD) is applied to the HRTF filter, based on the listener head circumference [11].

To achieve a good externalization of sound, distance and room reverberation effects should be simulated. Many simulators use artificial reverberation techniques to imitate the acoustic reverberation process [12]. Very few binaural reverb algorithms for real-time simulation exist that are based on a true-3D rendering, as standard mono or stereo reverberators are often used. Within the 3D Tune-In Toolkit, Binaural Room Impulse Response (BRIR) [13] are employed using a virtual-Ambisonic approach [14][15]. This consists in encoding sound sources in the Ambisonic domain, decoding then the Ambisonic channels in a set of speaker channels, which are finally 'converted' to binaural by convolving them with the BRIRs corresponding to their position. This method allows for a certain amount of flexibility (e.g. the complexity of the rendering can be reduced/increased by modifying the Ambisonic order and the number of speaker used for the decoding). Furthermore, thanks to Ambisonic encoding being a simple and 'light' operation, its computational weight is not as impacted by the number of spatialised sources as it would be when using direct convolution between each source and a BRIR.

Distance between listener and sound sources is simulated by a natural balance between the anechoic and reverb overall sound levels, and a modification of the Interaural Level Differences (ILD) for simulating sound sources closer than the position in which the HRTF has originally been measured. A similar approach has also been used in [16]. Finally, audio filters can simulate spectral changes related with air absorption at distances larger than 10-15 metres [17]. The 3D Toolkit implements the three mechanisms for distance simulation described above.

Only in Europe over 50 million people suffer from hearing loss [18]. Users with a hearing aid device are simply not able to wear a standard headphone. In order to fill the gap that exists between the people with hearing loss and the binaural sound experience, the 3DTI Toolkit includes a novel hearing aid simulator, which allows hearing impaired individuals to remove their hearing aid, wear a pair of headphones and compensate for their hearing loss using the virtual hearing aid simulator.

Finally, the Toolkit is completed with hearing loss simulation. This allows for development of applications aimed at enabling individuals with no hearing impairment to understand how hearing loss can compromise everyday activities, and how a hearing aid can improve this situation.

Loudspeaker-based 3D sound spatialisation is also well stablished [19]. In the 3D Tune-In project framework, 3D loudspeaker-based audio rendering is designed primarily for training hearing aid users. Naef [20] presents an overview loudspeaker-based spatialization techniques. The Loudspeakers audio renderer module implemented in the 3DTI-Toolkit uses the Ambisonic technique for spatial audio reproduction. Details of Ambisonic theory can be found in [21] and [22].

3 Related work

3D Audio spatialisation is a well stablished technology. However, to the best of our knowledge, there are only a few examples of free or open-source 3D audio rendering libraries that intend to implement a solution as comprehensive as the 3DTI Toolkit. One of them is the SoundScape Renderer (SSR) [23] which implements, among others, dynamic binaural synthesis and binaural room synthesis using HRTFs and BRIRs, respectively. However, SSR limits the source positions to the horizontal plane and it is currently available only for Linux and Mac OS X. In contrast, the 3DTI Toolkit offers full three-dimensional positioning of sources. At the moment of writing this paper, the 3DTI Toolkit has been tested on Windows, Mac OSX, Android, iOS and transpiled to Javascript using emscripten.

Other comparable library is the Steam Audio SDK [24], recently launched by Valve Corporation (before known as Phonom 3D audio SDK of Impulsonic, Inc). This SDK also implements a real-time HRTF-based binaural rendering and the sound propagation effects are applied as direction-dependent convolution reverb. Nevertheless, Steam Audio SDK uses a standardized HRTF, while the 3DTI Toolkit allows to use a customised HRTF.

Other free but less comprehensive solutions exist. Google Omnitone [25] is an open source library written in JavaScript for spatial audio processing. It features a first-order-Ambisonic decoder with HRTF binaural rendering. The Microsoft HRTF Spatialiser [26] implements HRTF binaural spatialisation for the HoloLens, based on DirectX and Unity. The Oculus Audio SDK [27], also features HRTF rendering to provide binaural audio spatialisation through the C/C++ SDK and plugins. This SDK allows developing for several platforms for the Oculus device, with the exception of the iOS platform. Oculus SDK also simulates room effects using a simple 'shoebox model'. In contrast, the 3DTI Toolkit provides binaural room synthesis. These last three libraries are free of charge to use but not distributed under an open source license.

Unlike the 3DTI Toolkit, none of the above software libraries provide, at the time of writing, hearing loss and aid simulation. Besides, only the SSR [23] features loudspeaker audio rendering using, among others methods, a Higher Order Ambisonic (HOA) simulation, but only circular setups can be used.

Commercial non open-source tools are also available. SPAT from IRCAM [28], which has been developed mainly for research/scientific use and works for both speakers and binaural rendering, is a real-time processor that allows the spatialisation of instrumental or synthetic sounds based on vector-based amplitude panning (VBAP) together with artificial reverberation. 3DCEPTION TwoBigEars [29], created for videogame developers, audio production and audio for films/videos, presents a spatial workstation for designing spatial audio for 360 video and cinematic VR.

There is a number of commercial hearing aid simulators, both for Android and iOS (e.g. Petralex hearing aid [30]). These simulators are configured automatically and do not present accessible low-level controls. Some very simplistic simulators for hearing loss can be found on the websites of hearing aid companies (e.g. Starkey [31]). These simulators cannot be used with custom audio content, in real-time and with simulation of non-linear features of the hearing loss.

4 Supported features

The 3DTI-Toolkit is organized into four main modules: binaural spatialisation, loudspeaker spatialisation, hearing loss and hearing aid simulators. The features of each module are described below.

Binaural spatialisation

Efficient Binaural Spatialisation of anechoic sound files is performed through convolution with Head Related Impulse Responses (HRIRs) interpolated to correspond with the desired source positions. The set of HRIRs (the HRTF) can be selected from a SOFA [32] file or a binary portable .3dti format developed within the 3D Tune-In project. The Toolkit can also add an extra shadow in the contralateral ear for sources very close to the listener's head. This process is based on the frequency-domain solution for the diffraction of an acoustic wave by a rigid sphere [33][34]. Furthermore, the ITD (Interaural Time Differences) can be re-computed to match that of a custom head circumference selected by the user. Distance simulation can also be performed for close and far sources. In addition to the anechoic spatialisation, the 3DTI Toolkit integrates binaural reverberation capabilities by convolving sources with room impulse responses. Using a novel approach based on a low-order Ambisonic encoding, reverberation is generated for all sources at the same time, but keeping certain location-dependent characteristics. This approach, together with an efficient convolution algorithm in the frequency domain, allows the Toolkit to compute large reverberating scenes, with virtually unlimited number of moving sources, maintaining high spatial accuracy for the direct sound (spatialized using direct-HRTF convolution). The listener head movements are also taken into account in the sound spatialisation.

Loudspeakers spatialisation

The 3DTI Toolkit can also perform loudspeaker-based sound spatialization. This has been implemented using the Ambisonic technique. Multiple sources are encoded in a 2nd Order Ambisonic stream, which is then decoded for various loudspeaker configurations, allowing the user to customise each speakers' position. The reverberation is generated using a similar approach to the one used for the binaural setup (i.e. based on virtual loudspeakers), allowing to simulate virtual environments in the Ambisonic domain with a fixed number of real-time convolutions, independently from the number of sources to be spatialised.

Hearing loss simulator

The 3DTI Toolkit can simulate different types of hearing loss. To configure the hearing loss parameters there are two possibilities for the users: (1) input a specific hearing loss curve, therefore inputting every value for each frequency band for the two ears independently, (2) simply select a preset, reporting whether a mild, moderate or severe hearing loss has to be simulated for each ear. The implementation of the hearing loss is, at the time of writing, in progress. This simulator will include frequency filters (e.g. graphic and dynamic equalisers), a dynamic range compressor/expander, non-linear distortion and degradation of temporal and spatial resolution. Currently, a basic hearing loss simulator is in place which implements already graphic equalization and dynamic range compressor.

Hearing aid Simulator

The Toolkit is able to simulate the user's hearing aid, whom will therefore not be required to wear any hearing aid devices when the auditory simulation is

headphones-based. The specific calibration of the virtual hearing aid, in terms of amplification and equalization, can be carried out individually for each user, either by entering the result of a standard audiological examination and using a widely used non lineal amplification algorithm called Fig6 [35], or configuring manually every parameter of the simulator. All parameters can be applied to each ear independently, therefore it would be beneficial to allow the user to control the left and right channels independently.
This simulator includes functions such as selective amplification, high/low pass filters, dynamic equalization (i.e. different equalization curves associated each to a different input audio level), directional processing (the directionality can be gradually selected from omnidirectional toward a "cardioid" pattern), dynamic range compression/expansion and re-quantisation (i.e. bitrate reduction).

5 Real-time Performance

The performance of the different modules of the 3DTI Toolkit for soft real-time applications has been measured, depending on the number of simultaneous audio sources and on the audio buffer size. All measures were obtained with an Intel Core i5-4440 computer with 8 GBs RAM running Microsoft Windows 7.

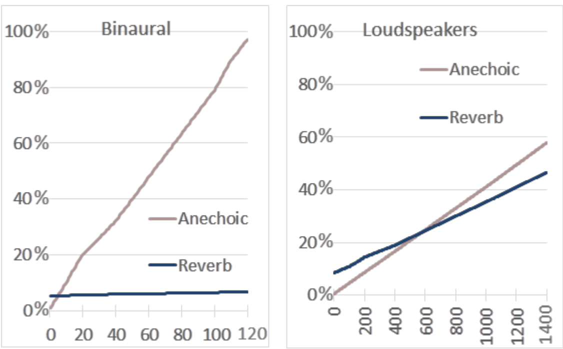

Figure 1. Performance of the spatialisers depending on the number of sources. The horizontal axis shows number of sources and the vertical axis shows percent of total buffer time (~11.61 ms). Left: binaural spatialiser. Right: loudspeakers-based spatialiser.

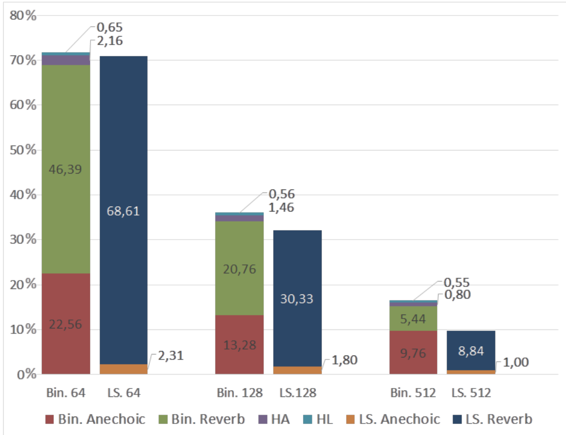

Figure 2. Performance of the Toolkit modules depending on buffer size. The vertical axis shows percent of total buffer time. The horizontal axis labels Bin.64, LS.64, correspond, respectively, to binaural processes with buffer size of 64 samples and loudspeakers-based processes with buffer size of 64; the same applies for the other labels, corresponding to buffer sizes of 128 and 512 samples for each group of processes.

Figure 1-Left shows how the performance of the binaural spatialiser processes scale with the number of sources. The vertical axis of the graph shows the percentage taken by each process from the total time available, when computing one buffer of 512 samples at a sample rate of 44100 Hz (~11.61 ms). The graph shows how the anechoic process (interpolated HRTF convolution and distance effects) cost has a growth linear with the number of sources, allowing up to 120 simultaneous sources before reaching the 100% of the available buffer time. The binaural reverb process cost is almost independent from the number of sources thanks to the Ambisonic approximation, where the only per-source process consists only on computing a 2nd order Ambisonic encoding. The BRIR used for this profiling has a size of 11345 samples. The performance of the hearing loss and hearing aid simulation processes is not shown in Figure 1-Left, because their cost is independent from the number of sources.

Figure 1-Right shows the performance of the loudspeakers-based spatialiser processes, depending on the number of sources. The vertical axis shows again the percentage of the available buffer time (~11.61 ms). In this case, both the anechoic and reverb processes grow slowly (and linearly) with the number of sources, since both per-source process compute a 2nd order Ambisonic encoding. The loudspeakers-based spatialiser supports up to 1400 sources with both anechoic and reverb processes. For profiling the reverb process, we used a set of 24 impulse responses, with 12406 samples each, where each file contained the impulse response for a given source position (north, south, east, west, zenith and nadir) and a given Ambisonic channel (W, X, Y, Z).

Figure 2 shows the performance of each module of the Toolkit relative to each other for different buffer sizes. All measures are taken for 10 simultaneous audio sources, a sample rate of 44100 Hz and the same impulse responses used for the other figures. The first two columns show the case of a very low latency buffer of 64 samples (~1.45 ms), demonstrating the capability of both the binaural (with hearing loss and hearing aid simulation) and loudspeakers-based spatialisers in high-demanding real-time applications. The cost is dominated by the reverb process, which depends only slightly on the buffer size thanks to the use of Uniformly-Partitioned Convolution [36]. For this reason, the proportion of time taken by the reverb process is halved each time the buffer size is doubled. This can be seen also for the binaural anechoic process, which implements Uniformly-Partitioned Convolution as well. The cost of the hearing loss and hearing aid simulation processes grow lineally with the buffer size. In the case of hearing aid simulation, the cost of the sub-processes which are independent on buffer size, dominates for small buffer sizes.

6 Conclusions

This paper presents the 3DTI Toolkit, an open-source C++ library developed within the EU-funded project 3D Tune-In.

This library enables developers to integrate efficient binaural and loudspeakers-based spatialisation into their applications. Binaural spatialisation is based on the Head Related Transfer Function, while loudspeaker spatialisation is based on Ambisonic. The 3DTI Toolkit features also convolution-based spatialised reverberation. In addition, and in contrast to comparable 3D audio rendering libraries, the 3DTI Toolkit allows developers to integrate in their application hearing loss and hearing aid simulation.

The library can be used directly via a C++ API. Additionally, a set of wrappers allow integration within other development environments including Unity (targeting Windows, Mac OSX, iOS and Android), Pure Data and JavaScript. At the time of writing, the development of the Toolkit was still active and had been tested in the aforementioned platforms with satisfactory performance.

A first version of the 3DTI Toolkit was planned for release with a non-commercial open-source license (royalty-free for non-commercial purposes, e.g. teaching and research) through the project website in Spring 2017. The full release was planned for release before May 2018.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644051.

References

[1] R. Eastgate, L. Picinali, H. Patel, and M. D'Cruz, "3D Games for Tuning and Learning About Hearing Aids," *Hear. J.*, vol. 69, no. 4, p. 30, (2016).

[2] Y. Levtov, L. Picinali, M. D'Cruz, and L. Simeone, "3D Tune-In: The Use of 3D Sound and Gamification to Aid Better Adoption of Hearing Aid Technologies." *Audio Engineering Society Convention 140*, (2016).

[3] Lord Rayleigh, "XII. On our perception of sound direction," *Philos. Mag. Ser. 6*, vol. 13, no. 74, pp. 214–232, (1907).

[4] P. Collins, "Theatrophone: the 19th-century iPod," *New Sci.*, vol. 197, no. 2638, pp. 44–45, (2008).

[5] J. Blauert, *The Technology of Binaural Listening*, (2013).

[6] V. Algazi and R. Duda, “Headphone-based spatial sound,” *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 33–42, (2011).

[7] “LISTEN HRTF DATABASE.” [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen/. [Accessed: 10-Mar-2017].

[8] “HRTF Measurements of a KEMAR Dummy-Head Microphone.” [Online]. Available: http://sound.media.mit.edu/resources/KEMAR.html. [Accessed: 10-Mar-2017].

[9] “The CIPIC HRTF Database - CIPIC International Laboratory.” [Online]. Available: http://interface.cipic.ucdavis.edu/sound/hrtf.html. [Accessed: 10-Mar-2017].

[10] B. F. G. Katz and G. Parseihian, “Perceptually based head-related transfer function database optimization.,” *J. Acoust. Soc. Am.*, vol. 131, no. 2, p. EL99-105, 2012.

[11] R. O. Duda, “Modeling head related transfer functions,” in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, 1993, pp. 996–1000, (1993).

[12] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, “Fifty Years of Artificial Reverberation,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 20, no. 5, pp. 1421–1448, (2012).

[13] D. N. Zotkin, R. Duraiswami, and L. S. Davis, “Rendering Localized Spatial Audio in a Virtual Auditory Space,” *IEEE Trans. Multimed.*, vol. 6, no. 4, pp. 553–564 (2004).

[14] L. Picinali, A. Afonso, M. Denis, and B. F. G. Katz, “Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge,” *Int. J. Hum. Comput. Stud.*, vol. 72, no. 4, pp. 393–407, (2014).

[15] A. McKeag and D. S. McGrath, “Sound Field Format to Binaural Decoder with Head Tracking.” 101st AES Convention, Los Angeles (CA), (1996).

[16] Katz, B., E. Rio, and L. Picinali. "LIMSI Spatialization Engine." *Inter Deposit Digital Number:* F 1.000 (2010).

[17] A. D. Little, D. H. Mershon, and P. H. Cox, “Spectral content as a cue to perceived auditory distance.,” *Perception*, vol. 21, no. 3, pp. 405–416, (1992).

[18] European Federation of Hard of Hearing People, “State of subtitling access in EU 2011 Report Subtitling in the European Union,” (2011).

[19] T. Lentz, D. Sch Oder, M. Vorï Ander, and I. Assenmacher, “Virtual Reality System with Integrated Sound Field Simulation and Reproduction,” *EURASIP J. Adv. Signal Process.*, vol. 70540, no. 19, (2007).

[20] M. Naef, O. Staadt, and M. Gross, “Spatialized audio rendering for immersive virtual environments,” in *Proceedings of the ACM symposium on Virtual reality software and technology - VRST ’02*, p. 65, (2002)

[21] M. A. Gerzon, “Ambisonics in Multichannel Broadcasting and Video,” *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, (1985).

[22] D. G. Malham and A. Myatt, “3-D Sound Spatialization using Ambisonic Techniques,” *Comput. Music J.*, vol. 19, no. 4, p. 58, (1995).

[23] M. Geier and S. Spors, “Spatial Audio with the SoundScape Renderer,” *27th Tonmeistertagung - VDT Int. Conv.*, (2012).

[24] “Steam Audio.” [Online]. Available: https://valvesoftware.github.io/steam-audio/. [Accessed: 10-Mar-2017].

[25] Google, “OMNITONE.” [Online]. Available: http://googlechrome.github.io/omnitone/#home. [Accessed: 13-Mar-2017].

[26] Microsoft, “Spatial sound.” [Online]. Available: https://developer.microsoft.com/es-es/windows/holographic/spatial_sound. [Accessed: 13-Mar-2017].

[27] Oculus VR, “Centro para desarrolladores | Oculus.” [Online]. Available: https://developer.oculus.com/. [Accessed: 13-Mar-2017].

[28] “IRCAM Forumnet | Spat.” [Online]. Available: http://forumnet.ircam.fr/product/spat-en/. [Accessed: 10-Mar-2017].

[29] “Two Big Ears | 3Dception.” [Online]. Available: http://www.twobigears.com/3dception.php. [Accessed: 10-Mar-2017].

[30] “PETRALEX — Hear It Clear!” [Online]. Available: http://petralex.pro/es. [Accessed: 10-Mar-2017].

[31] “Hearing Loss Simulator - Find Out What Hearing Loss is Like.” [Online]. Available: http://www.starkey.co.uk/hearing-loss-simulator. [Accessed: 10-Mar-2017].

[32] “SOFA (Spatially Oriented Format for Acoustics) - Sofaconventions.” [Online]. Available: https://www.sofaconventions.org/mediawiki/index.php/Main_Page. [Accessed: 10-Mar-2017].

[33] L. Rayleigh and A. Lodge, “On the Acoustic Shadow of a Sphere,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 203, no. 359–371, pp. 87–110, Jan. (1904).

[34] J. W. S. Rayleigh and R. B. Lindsay, *The theory of sound*. Dover, (1945).

[35] T. C. Gitles and P. Tillman Niquette, “FIG6 In Ten.” Hear Rev, 2(10), 28-30 (1995).

[36] F. Wefers, *Partitioned convolution algorithms for real-time auralization*, vol. 20, (2015).