ANALYZING NOTCH PATTERNS OF HEAD RELATED TRANSFER FUNCTIONS IN CIPIC AND SYMARE DATABASES

M. Shahnawaz, L. Bianchi, A. Sarti, S. Tubaro

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy

ABSTRACT

The sensation of elevation in binaural audio is known to be strongly correlated to spectral peaks and notches in HRTFs, introduced by pinna reflections. In this work we provide an analysis methodology that helps us to explore the relationship between notch frequencies and elevation angles in the median plane. In particular, we extract the portion of the HRTF due to the presence of the pinna and we use it to extract the notch frequencies for all the subjects and for all the considered directions. The extracted notch frequencies are then clustered using the K-means algorithm to reveal the relationship between notch frequencies and elevation angles. We present the results of the proposed analysis methodology for all the subjects in the CIPIC and SYMARE HRTFs databases.

Index Terms— Binaural audio, Elevation perception, Head Related Transfer Function (HRTF), k-means.

1. INTRODUCTION

Sound perception is the result of the interaction between the acoustic wavefield and the listener's body, which causes wave scattering, reflection and diffraction. These phenomena alter the spectral content of the sound signal in a direction-dependent fashion, and introduce a wide variety of cues that enable sound localization. The interaction between soundfield and listener's body is encoded by a complex-valued transfer function, usually known as *Head Related Transfer Function (HRTF)*, which describes the spectral modifications that are characteristics of a source in a given location with respect to the listener [1]. The time-domain equivalent of this transfer function is known as *Head Related Impulse Response (HRIR)*.

Knowing the HRTF of a person is what enables spatial sound reproduction using headphones. However, as confirmed by many studies, HRTFs are strongly dependent on the listener's anatomy. This means that, in order to guarantee the best performance in terms of sound localization, individualized HRTFs need to be adopted [2, 3]. Unfortunately, the measurement of HRTFs is so expensive and time-consuming to prevent its use in consumer applications.

A great deal of effort has been put into the personalization of HRTFs. In [4, 5], for example, suggest to estimate individualized HRTFs from 3D models of the user's pinnas. Some techniques based on low-cost capturing devices [6] have been proposed for this purpose, though the acquisition of a sufficiently accurate 3D model is still not an easy task for the average user. An alternate solution consists of synthesizing individualized HRTFs from a structural model of the listener's body [7–9]. Using parametric filters that rely on a given mapping between parameters and anthropometric data, the authors obtain computationally efficient and customizable solutions that can be used to approximate individualized HRTFs.

Notches in the HRTF caused by the pinna, are known to have significant perceptual relevance for sound localization, particularly in the frontal region [10–13]. Some studies, e.g. [14, 15], reported that the frequencies of the notches greatly depend on the elevation angle of the sound source, and they are almost independent of azimuth and distance. Recently, an important observation has been made in [4], where the authors related the notches in the HRTF with the three main pinna contours.

In this manuscript we study the relation between the notch frequencies and the elevation angles for a large number of subjects, whose HRTFs have been acoustically measured and stored into two databases: the CIPIC database [16] and the SYMARE database [5]. Notch frequencies are extracted from the collected HRTFs after removing all contributions of head, torso, and shoulders, while retaining only the contribution of the pinna, as described in [17]. We group the notch frequencies for all the subjects under consideration into three clusters, each corresponding to one of the three main pinna contours identified in [4]. In this setting, we analyze the evolution of the notch frequencies as a function of sound source elevation in the median plane. Moreover, we analyze the correlation between notch frequencies in the left and right ears.

2. ANALYSIS METHODOLOGY

This section introduces the methodology used in this work. The overall methodology can be divided into two conceptual steps: notch frequency extraction and clustering. Figure 1 shows the block diagram of the overall analysis methodology while Fig. 2 explains the steps involved in notch extraction.

We would like to thank Prof. Craig Jin and Dr. Nicolas Epain and CAR-Lab research team for all the help and support, as well as for the permission to use the SYMARE database for our work.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

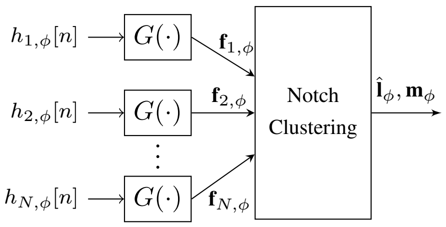


Fig. 1 Block diagram of the proposed analysis methodology.

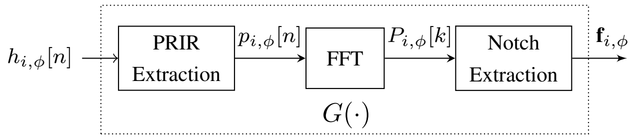


Fig. 2 Detail of the notch extraction procedure.

The detailed description of each step is given below.

2.1. PRTF Extraction

The deep spectral notches are produced in HRTF due to reflections caused by different body parts including pinna cavities, head, torso and knees. In this study we aim to analyze the spectral notches caused by pinna, so the first step is removing all unnecessary components of HRIR namely the contributions of head, shoulders and knees preserving the contributions of pinna. In [17] it was reported that the delays of pinna, torso and knee reflections are typically around 0.1 to 0.3, 1.6 and 3.2 ms [10, 17] respectively.

To get rid of shoulders, torso and knees reflection components we shorten our HRIR by applying a half Hanning window [17] of length 1 ms, starting from onset of HRIR. This removes the reflective components due to shoulders, torso and knees, while preserving the reflection caused by pinna.

Given the HRIR $h_{i,\phi}[n]$ for the user i, and elevation angle ϕ, the PRIR $p_{i,\phi}[n]$ can be extracted by applying a half Hanning window $w[n]$ starting from onset of HRIR n_o, i.e. $p_{i,\phi}[n] = h_{i,\phi}[n]w[n - n_0]$. Figure 3 illustrates the windowing operation. The value of n_0 can be found by taking the slope of unwrapped phase function of HRTF [17].

Once the PRIR $p_{i,\phi}[n]$ is obtained, the PRTF (Pinna related transfer function) $P_{i,\phi}[f]$ can be obtained by evaluating its Fourier transform, where f denotes the frequency of the signal. Next we describe the notch frequency extraction procedure from the PRTFs $P_{i,\phi}[f]$, $i = 1, \ldots, N$, relative to all N users.

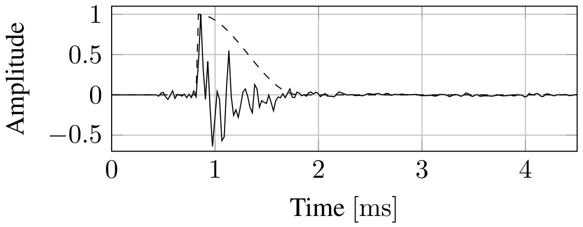


Fig. 3 HRIR windowing for PRIR extraction.

2.2. Notch Extraction

As reported in [11], the frequency content in the range 4 kHz to 16 kHz is the main cause of median plane localization. For this reason, we restrict the frequency bandwidth of our analysis to this range.

To extract the notches we use the negative of log-scale magnitude function of the PRTFs, i.e.

$$\mathring{P}_{i,\phi}[f] = -20\log_{10}(|P_{i,\phi}[f]|). \tag{1}$$

The purpose of this step is to turn the notches into peaks, so that they can be effectively extracted by finding the local maxima in $\mathring{P}_{i,\phi}[f]$. In order to get meaningful results, we also have to make sure that we are considering just the significant and prominent notches, while discarding all those which are not relevant. For this purpose, we consider the prominence of the local maxima. The prominence describes how much the peak stands out from the neighboring peaks. For instance, a low isolated peak can be more prominent than one that is higher but is next to an other higher peak and vice-versa.

In the following, we considered those peaks in $\mathring{P}_{i,\phi}[f]$ that have a prominence greater than 3 dB. These values are stored in vectors $\mathbf{f}_{i,\phi}$ for each subject i and elevation ϕ as

$$\mathbf{f}_{i,\phi} = [f_{i,\phi,1}, \ldots, f_{i,\phi,M_{i,\phi}}], \tag{2}$$

being $M_{i,\phi}$ the number of relevant peaks in PRTF of i^{th} user for elevation angle ϕ.

Once we have notch frequency vectors, $\mathbf{f}_{i,\phi} \in \mathbb{R}^{1\times M_{i,\phi}}$ for all the users and elevations, we arrange them into the vector $\mathbf{f}_\phi$, which contains the notch frequencies for all the users for elevation ϕ, i.e.

$$\mathbf{f}_\phi = \left[\mathbf{f}_{1,\phi}\mathbf{f}_{2,\phi}\ldots\mathbf{f}_{N,\phi}\right] \in \mathbb{R}^{1\times M_\phi}, \text{ with } M_\phi = \sum_{i=1}^{N} M_{i,\phi}. \tag{3}$$

2.3. Clustering of Notches

The next step of the analysis is to find the meaningful information from the frequency vectors $\mathbf{f}_\phi$. In a recent study [4], the authors reported that in each PRTF in CIPIC database up to three main spectral notches can be extracted, and mapped to three distinctive and prominent pinna contours: the helix, anti helix and outer wall of the concha.

Based on these findings, in this study we clustered the notch frequency vector $\mathbf{f}_\phi$ consisting of M_ϕ elements into $K = 3$ groups, using a well known clustering algorithm K-means [18]. At the end of the process, each element in $\mathbf{f}_\phi$ will be assigned to a single cluster, whose centroid is the closest to the actual value of the element.

We evaluate the distance between each element $f_{i,\phi,j} \in \mathbf{f}_\phi$ and the corresponding centroid $m_{k,\phi}$ as the euclidean distance $D(f_{1,\phi,j}, m_{k,\phi}) = |f_{1,\phi,j} - m_{k,\phi}|$.

The K-means algorithm is initialized by assigning random values to the centroids $m_{k,\phi}$, $k = 1, 2, 3$. The algorithm is defined as an iterative two-step process. The first step is the assignment of each notch frequency to a cluster having closest centroid and label it with the number of that cluster e.g. $1, 2$ or 3 according to

$$\hat{l}_j = \arg\min_k \{D(f_{i,\phi,j}, m_{k,\phi})\} \tag{4}$$

where $j = 1, \ldots, M_\phi$ and $k = 1, 2, 3$. Moreover, a responsibility vector is defined for each cluster as

$$r_{k,j} = \begin{cases} 1, & \text{if } \hat{l}_j = k, \\ 0, & otherwise. \end{cases} \tag{5}$$

The second step is to update the centroid for all the clusters. The updated value for the kth centroid is

$$m_{k,\phi} = \frac{\sum_{j=1}^{M_\phi} r_{k,j} f_{i,\phi,j}}{R_k}, \quad R_k = \sum_{j=1}^{M_\phi} r_{k,j} \tag{6}$$

where R_k is the total responsibility of cluster k, defined as the number of points belonging to cluster k.

The process continues until no further changes occur in the cluster centroids.

After applying the K-means algorithm, we obtain the centroids $\mathbf{m}_\phi = [m_{1,\phi}, m_{2,\phi}, m_{3,\phi}]$, corresponding to helix, anti-helix and outer wall of concha respectively. Moreover, in order to associate a relevance descriptor to the clustered data, we introduce the cluster spread as the standard deviation of their elements, i.e.

$$\sigma_{k,\phi} = \sqrt{\frac{\sum_{j=1}^{M_\phi} (f_{i,\phi,j} - m_{k,\phi})^2 \, r_{k,j}}{R_k}} \tag{7}$$

The results are further analyzed in the next section.

3. RESULTS

In this section we describe the application of the analysis methodology described in Sec. 2 to the CIPIC and SYMARE databases.

3.1. Description of the databases

For this study we used acoustically measured HRTFs from two well known databases of fairly large population set.

3.1.1. CIPIC

CIPIC [16] is a public-domain database of acoustically measured HRIRs with a high spatial resolution. It contains HRIRs for 45 subjects (27 male, 16 female and two KEMAR) measured at 1250 different directions around the head of the subjects. The measurements are done using Golay code as analysis signals, with a sampling frequency of $44.1\,\mathrm{kHz}$. Measurements loudspeakers are mounted on a circular arc of radius $1\,\mathrm{m}$, which is rotated around a fixed listener. The length of each HRIR stored in the database is 200 samples. For the purpose of this work, we consider all the HRIRs at azimuth 0° and elevations ϕ between $-45°$ and $45°$, with a uniform spacing equal to $5.625°$.

3.1.2. SYMARE

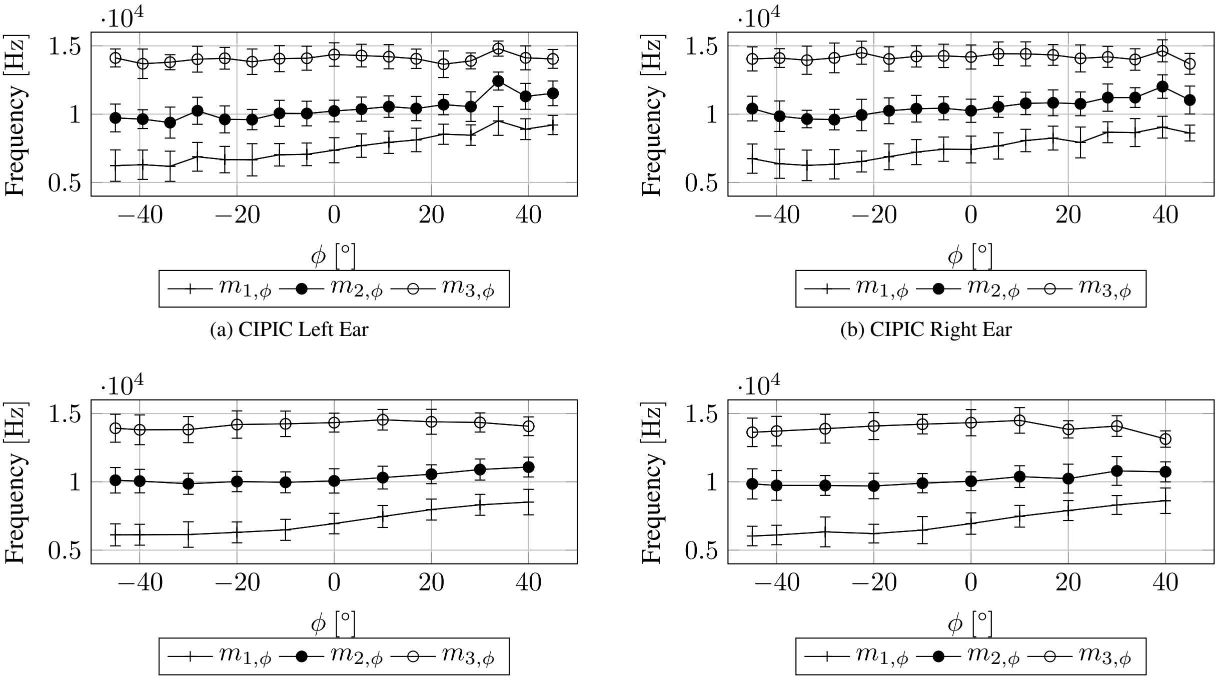
SYMARE [5] database was created by a collaborative team of Sydney University Australia and University of York England. This database contains acoustically measured HRTFs for 61 users (45 males and 16 females) measured in 393 directions around the head at a distance of $1\,\mathrm{m}$, with a non-uniform angular spacing in elevation for different azimuth angles. Impulse responses are recorded using Golay codes with a sampling frequency equal to $48\,\mathrm{kHz}$. The length of each HRIR is 256 samples. For the purpose of this work, we consider all the HRIRs at azimuth 0° and elevations ϕ between $-45°$ and 40°.

3.2. Analysis 1

The steps defined in section 2 were applied to all the HRIR sets in both databases. HRIRs for the mentioned elevations were retrieved from the databases and PRIRs were extracted from each HRIR. The PRIRs were then transformed in the frequency domain by a zero-padded 512-point FFT.

Notches vectors $\mathbf{f}_\phi$ are estimated for each direction ϕ according to the angular grid adopted by the database, and notch frequencies are grouped into 3 clusters $m_{k,\phi}$, $k = 1, 2, 3$, along with their corresponding spread $\sigma_{k,\phi}$. Figure 4 shows the cluster centroids and spreads as a function of the elevation angle ϕ for the left and right ears of all the subjects in CIPIC and SYMARE databases.

We notice that for $\phi = -45°$ the cluster mean for all four cases (CIPIC and SYMARE databases, left and right ears) has almost the same value. Another observation that we want to point out is that all the cluster means $m_{k,\phi}$, $k = 1, 2$, exhibit a monotonically increasing behavior as a function of ϕ, despite of some slight irregularities. These irregularities are more prominent in the CIPIC database. On the other hand, $m_{3,\phi}$ results to be almost constant in all the four considered cases. In a more general way, we observe that the slope of the clusters $m_{k,\phi}$, $k = 1, 2, 3$, is the highest for $m_{1,\phi}$ and almost null for $m_{3,\phi}$. This behavior suggests that the pinna reflection causing a notch in the range of $m_{1,\phi}$ might be the most



(a) CIPIC Left Ear

(b) CIPIC Right Ear

(c) SYMARE Left Ear

(d) SYMARE Right Ear

Fig. 4 Cluster centroids and spreads as a function of elevation angle ϕ.

informative one for elevation perception.

In the case of data extracted from the CIPIC database, we observe a peak around $\phi = 30°$ for the left ear, while the right ear exhibit a peak around $\phi = 40°$. In the SYMARE database these irregularities are very mild and are present in just right ear, while the tracks for left ear are very smooth.

3.3. Analysis 2

Further, we compare the results obtained for left and right ears in both databases. First, we convert the frequency centroids $m_{k,\phi}$ to the Bark scale [19] and then we compute their Euclidean distance. In the following we denote by $d_{k,\phi}$ the distance between the centroid of left and right ears for the kth cluster and elevation ϕ. Results are reported in Fig. 5.

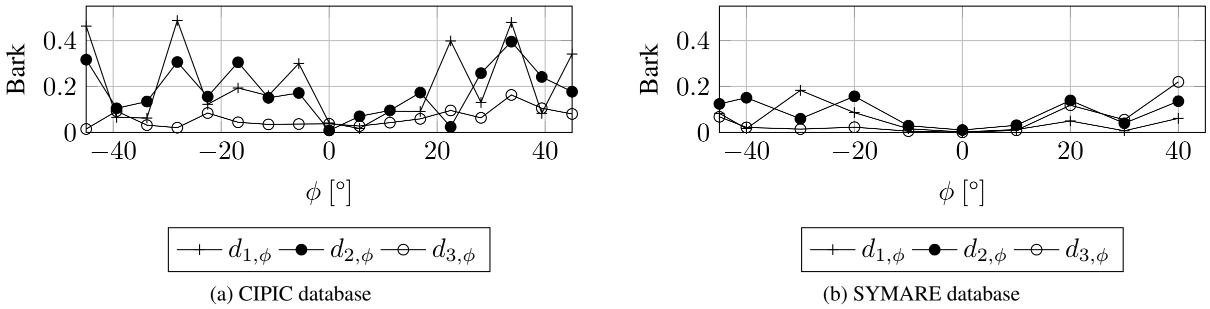
We observe that, in both CIPIC and SYMARE, the maximum value for the distance between clusters is less than 0.5 Bark for all the considered cases and for all the elevations. In case of SYMARE database distances have smaller values and a smoother distribution, while in the CIPIC database distances are, in general, greater and less regular with respect to ϕ. We would like to point out that the differences exhibit minima in the horizontal plane ($\phi = 0°$) in all the considered cases and for all the clusters, suggesting that binaural cues are not relevant in the frontal direction. On the other hand, it can be observed that the distances are greater moving away from the horizontal plane; this behavior suggests that both monaural and binaural cues are relevant for elevation perception in the median plane.

4. CONCLUSIONS

In this manuscript we provide a methodology to analyze HRTFs in publicly available databases. In particular, we describe a technique to extract notch frequencies from HRTF data and to classify them into three clusters, each corresponding to a specific contour in the pinna namely the helix, anti helix and outer wall of the concha. We validated the proposed methodology with acoustically measured HRTFs from the CIPIC and SYMARE databases. We performed a comparative study on the evolution of notch frequencies in median plane in CIPIC and SYMARE databases. Results show the strong dependency between notches in the HRTFs and elevation angles in the median plane. Moreover, we also studied the binaural differences between noth frequencies which revealed that not only monaural but also binaural cues are important for elevation perception. We envision our approach to be applied in combination with the techniques mentioned in [20–22] for the auralization of virtual and real sound environments.

REFERENCES

[1] C. I. Cheng and G. H. Wakefield, “Introduction to head-related transfer functions (HRTFs): Representations of HRTFs



(a) CIPIC database

(b) SYMARE database

Fig. 5 Distance between the centroids for the left and right ears as a function of elevation ϕ.

in time, frequency, and space," in *Proc. AES 107th Conv.* AES, 1999.

[2] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[3] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.

[4] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the relation between pinna reflection patterns and Head-Related Transfer Function features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 3, pp. 508–519, 2013.

[5] C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the Sydney York morphological and acoustic recordings of ears database," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 37–46, 2014.

[6] L. Bonacina, A. Canclini, F. Antonacci, M. Marcon, A. Sarti, and S. Tubaro, "A low-cost solution to 3D pinna modeling for HRTF prediction," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2016.

[7] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.

[8] V. R. Algazi, R. O. Duda, and P. Satarzadeh, "Physical and filter pinna models based on anthropometry," in *Proc. AES 122nd Conv.* AES, 2007.

[9] I. Faller, K. John, A. Barreto, and M. Adjouadi, "Augmented Hankel total least-squares decomposition of head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 58, no. 1/2, pp. 3–21, 2010.

[10] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 168, no. 1011, pp. 158–180, 1967.

[11] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.

[12] D. Wright, J. H. Hebrank, and B. Wilson, "Pinna reflections as cues for localization," *Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 957–962, 1974.

[13] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, vol. 68, no. 8, pp. 835–850, 2007.

[14] E. A. G. Shaw, *Binaural and Spatial Hearing in Real and Virtual Environments*, chapter Acoustical features of the human external ear, Lawrence Erlbaum, Mahwah, NJ, US, 1997.

[15] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, 1999.

[16] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2001.

[17] V. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, 2005.

[18] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[19] H. Traunmüller, "Analytical expression for the tonotopic sensory scale," *Journal of the Acoustical Society of America*, vol. 88, pp. 97–100, 1990.

[20] M. Foco, P. Polotti, A. Sarti, and S. Tubaro, "Sound spatialization based on fast beam tracing in dual space," in *Proc. of the 6^{th} Int. Conference on Digital Audio Effects, (DAFx-03)*, 2003.

[21] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Real time modeling of acoustic propagation in complex environments," in *Proc. of the 7^{th} Int. Conference on Digital Audio Effects DAFx-04*, 2004.

[22] M. Vorländer, *Auralization, Fundamentals of acoustics, modemodel, simulations, algorithms and acoustic virtual reality*, Springer-Verlag Berlin Heidelberg, 2008.