<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>On the study of frontal-emitter headphone to improve 3D audio playback</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Sunder, Kaushik; Tan, Ee-Leng; Gan, Woon-Seng</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Sunder, K., Tan, E.-L., &amp; Gan, W.-S. (2012). On the study of frontal-emitter headphone to improve 3D audio playback. 133rd Audio Engineering Society Convention 2012.</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/26353">http://hdl.handle.net/10220/26353</a></td>
</tr>
</tbody>
</table>

© 2012 Audio Engineering Society (AES). This paper was published in 133rd Audio Engineering Society Convention 2012 and is made available as an electronic reprint (preprint) with permission of Audio Engineering Society (AES). The published version is available at: [http://www.aes.org.ezlibproxy1.ntu.edu.sg/elibri/inst/browse.cfm?elib=16502]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.
On the study of frontal-emitter headphone to improve 3D audio playback

Kaushik Sunder 1, Ee-Leng Tan 1, and Woon-Seng Gan 1

1 DSP Lab, School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, 639798
{kaushik1, etanel, ewsgan}@ntu.edu.sg

ABSTRACT

Virtual audio synthesis and playback through headphones by its virtue have several limitations, such as the front-back confusion and in-head localization of the sound presented to the listener. Use of non-individual head related transfer functions (HRTFs) further increases front-back confusion and degrades the virtual auditory image. In this paper, we present a method for customizing non-individual HRTFs by embedding personal cues using the distinctive morphology of the individual’s ear. In this paper, we study the frontal projection of sound using headphones to reduce the front-back confusion in 3D audio playback. Additional processing blocks, such as decorrelation and front-back biasing are implemented to externalize and control the auditory depth of the frontal image. Subjective tests are conducted using these processing blocks and its impact to localization is reported in this paper.

1. INTRODUCTION

One of the biggest challenges in 3D audio is to reproduce the spatial information with conventional headphones as faithfully as possible. The transmission of sound from a point in space to the ear drum of the listener is encoded in the form of acoustical transfer functions, also known as the head related transfer functions (HRTF) [1]. Binaural audio can be synthesized by filtering the original sound source with a pair of HRTFs corresponding to the location of the desired virtual auditory image in the physical space. 3D audio playback through headphones is the most preferred method since it allows independent control of the acoustic pressure exciting the two channels and avoids crosstalk. But this playback mode comes with other inherent deficiencies, such as localization confusions and coloration of the source spectrum. A highly accurate 3D audio synthesis is possible only when we have individualized HRTFs and individualized equalization, which are tedious to measure [2]. The impracticality of measuring individual HRTFs for every human subject leads to the development of a generalized artificial human head (dummy head), which simplifies the process of HRTF measurement. The dummy head is constructed in a way so as to include all the important
features of a human upper body, which affects sound localization. As a result, researchers have made several HRTF databases freely available in the public domain [3, 4]. However, the use of such non-individualized HRTFs leads to the degradation of the spatial auditory display. Thus, one of the current research problems is to enhance the spatial auditory display using non-individual HRTFs.

Principle component analysis (PCA) methods have been devised to model individual HRTFs as a weighted sum of basis functions, which are common to all the individuals [5]. Besides this model based approach, other simpler methods to improve the non-individualized spatial auditory display are to ask the listener to choose a HRTF from the database that best fits his or her perception [6]. Another method is to ask the subject to tune the magnitude spectrum within the frequency bands in order to customize the non-individualized HRTFs [7]. These methods require some forms of training or queries to enhance the individual cues.

In this paper, we study the frontal projection of sound (Figure 1), which would emulate the acoustical interactions at high frequencies with the pinna and the concha, thereby mimicking the sound projection from a pair of physical loudspeakers. We study the characteristics of such frontal emitters and differentiate it with the conventional side emitter headphones. Similar work is done in the past by Tan et al. [8], who studied the direct concha excitation of the sound. The work reported in this paper extends on the latter work by introducing additional processing blocks, such as the decorrelation and front-back biasing to further enhance the spatial auditory effect. We further develop a method to use the front back biasing to control the auditory depth of the frontal image created. In addition, we also present a technique of equalizing the frontal emitter, which differs from equalizing the conventional side emitter headphones. The theoretical analysis is later corroborated by the subjective experiments where we report the effect of equalization and also the contribution of each processing block.

This paper is organized as follows: In Section 2, we describe the role of the pinna in localization and the spectral cues associated with the frontal directions. Description of the measurements and its objective analysis are presented in Sections 3, 4 and 5 respectively. The subjective tests and results are described in Section 7.

2. RELATED BACKGROUND

Human beings make use of three fundamental localization cues: Interaural Time Differences (ITD), Interaural Level Differences (ILD) and Spectral Cues (SC) to identify the direction of the sound. ITD and ILD are the principal cues for sound localization based on the duplex theory proposed by Rayleigh in 1907. ITD and ILD mainly cue the localization to the left or right based on the time and level differences between the two ears. ITD is an important cue at low frequencies (below 1500 Hz) and ILD operates mainly at high frequencies (above 1500 Hz) due to the head shadowing effect. SC arises as a result of reflections and diffractions of sound with the external ear. Human head, torso and shoulders also play a significant role in modifying the spectrum of the sound source at the eardrum. All these above mentioned cues are encoded by the head related impulse response (HRIR). The pinna plays a critical role in frontal localization of sound source. In the following subsections, we shall describe the spectral cues caused by the pinna interaction and its importance in localization, in particular the frontal localization.

2.1. Role of Pinna

The pinna cues are the most widely studied amongst all the other localization cues. Hebrank and Wright [9] observed a unique spectral notch generated by the interference of direct sound entering the external auditory canal and the time-delayed reflections off the posterior concha wall. They carried out localization experiments with high-pass, low-pass, band-pass and band-stop filtered white noise as stimulus and observed that the sound spectra from 4 to 16 kHz are necessary for sound localization. This observation implies that the sound source must be broadband for the listener to localize accurately. The directional cues for the frontal,
overhead and rear directions as reported in [9] are summarized in Table 1. Frontal directions are mainly cued by a 1 octave BW notch above 4 kHz [9]. An increase in elevation in this direction is cued by an increase in the lower cut-off frequency of the 1 octave BW notch mentioned above [10]. Other psychoacoustical studies have also demonstrated that the high frequency pinna cues are necessary for accurate localization [9, 11].

Several studies have been carried out to understand the influence of the anatomical features of the external ear. Pinna occlusion experiments conducted by Gardner et al. [11] showed that the localization ability decreases with increasing occlusion. They found that each of the three main parts of the pinna: Scapha, Fossa and Concha are important for accurate localization. The occlusion of concha degraded the localization performance the most. Hofman et al. [12] changed the shape of the ears by inserting plastic molds in the pinna cavity and measured the localization performance. Elevation localization was severely degraded immediately after the modification, however accurate localization was acquired again after some training. All the above experiments indicate that pinna plays a critical role in localization of sound source in the real auditory space.

### 2.2. Localization performance in real and virtual space

Wightman and Kistler [13] performed several localization experiments with real, as well as virtual sound sources to examine front-back confusion. With head movements restricted, 3 out of 7 subjects experienced confusion while localizing a real source. For the same listening test, subjects experienced very few front-back confusions when head movements are permitted.

Localization of virtual sound sources using individualized HRTFs degraded slightly by 1 or 2 degrees in comparison with real sources. The elevation perception was the most affected and the front-back confusions were almost doubled. With head tracking, the front-back confusions were reduced except for a few subjects, but the elevation perception still remained the most affected. However, the virtual sound sources were well externalized in this case.

Wenzel et al. [14] conducted a detailed study on the localization using non-individual HRTFs. Their studies suggested that localization performance with non-individual HRTFs depended highly on the individual, but the localization was degraded for most of the subjects. The azimuth judgment was very much accurate, while the elevation and frontal localization was the most affected in this case. The externalization of the frontal sound source was poor and was most likely perceived inside the head. In addition, the sound sources appeared to be blurred when non-individual HRTFs were used. Use of dynamic HRTFs with the help of head tracking, which needs very high processing power, improved the localization performance but the confusions were still not completely resolved. This is because the pinna is one of the most individual elements of the morphology of the listener's ear and thus, the individual spectral cues are highly disturbed by the use of the non-individual HRTFs. Frontal projection of the sound recreates these missing individual spectral cues in the non-individual HRTFs and aids in resolving the ambiguity in localization in the region of cone of confusion.

### 2.3. Frontal Localization

As described earlier, one of the major drawbacks in conventional headphones playback using non-individual HRTFs is the front-back confusion. Blauert [1] mentioned that the correct frontal perception is the result of a combination of a proper loudness balance between the different frequency bands. Researchers have also tried to amplify the spectral difference between the front and rear directions to enhance the perceptual differences between the sounds coming from these directions [15]. Frontal direction HRTFs depict a unique spectral notch created by the interactions with the external ear, which displays large inter-individual variability as shown in Figure 2. Poveda and Meddis [10] developed a physical model of diffractions and reflections in the human concha. They also provided evidence indicating that, similar spectral
features were generated in the concha for sources at all azimuths within the frontal part of the ipsilateral hemisphere (Figure 2).

From Figure 2, we see that the spectral notch frequency is almost constant in the frontal directions for a given subject. Since the position of the frontal emitter is maintained fixed relative to the ear, the notch created by the frontal projection of the sound does not vary for a particular individual. This fact coincides well with the above observation, and thus the frontal projection can be used to recreate cues throughout the frontal azimuthal plane. The following section describes the procedures for measurement of frequency response of the frontal and the side emitter headphones, followed by the objective analysis of the results in the subsequent section.

3. **MEASUREMENTS**

The frequency responses of the frontal and the side emitters were measured to carry out the objective analysis of the frontal and side projection over a headphone. These measurements were later used to equalize the emitters. The impulse responses were measured on the B&K Head and Torso simulator. The measurements were conducted in a semi-anechoic chamber at the DSP lab in NTU. Exponential sine sweep signals were used as stimulus to measure impulse responses of the headphones. This signal was preferred over pseudo noise signals (MLS signal) because they show higher immunity against distortion and time variance. Higher SNR levels (greater than 90 dB) can be achieved using sweeps due to the possibility of complete rejection of harmonic distortion, which is unattainable with MLS measurements [16]. Special care was taken to ensure that sine sweep is long enough to allow the complete rejection of the harmonic distortion, and to ensure that the low frequency energy is present for sufficient amount of time. The exponential sine sweep was sent through the B&K PULSE system to the emitters and the PULSE system captured the recorded signal at the eardrum of the dummy head (Figure 6). The impulse response was extracted from the recorded data in MATLAB. During the measurement process, the headphone is removed and repositioned on the dummy head after each trial and 10 such trials are performed. The mean transfer function of all the trials is computed across 10 trials.
frontal emitters are plotted in Figure 3 and Figure 4, respectively.

The free field frequency response (FFR) of only the emitter along with the ear cup was also measured for the frontal emitter and is plotted in Figure 5. This FFR is used to equalize the frontal emitter headphone. This process will be explained in detail in Section 5. Unlike the transfer function measured at the eardrum of the dummy head, this response does not include the headphone-ear coupling. The standard deviation of the different trials is also shown at the bottom (in black) of Figures 3 - 5. It is found that the variation is larger at the higher frequencies as compared to the lower frequencies. Kulkarni et al. [18] explained that this observation is due to the similar pressure within the earphone cavity at low frequencies.

4. OBJECTIVE ANALYSIS

The mean values of the frontal emitter response (FER) and the side emitter response (SER) are overlaid together and plotted in Figure 7. A closer look at this plot reveals that the low-frequency response of the frontal emitter is 6 dB lower compared to the side emitters. This depicts an improvement in the low-frequency response of the frontal emitter compared to the 20 dB drop as reported by Tan et al. [8]. The peak around 2-4 kHz in both FER and SER is due to ¼ wavelength ear canal resonance. The notch at 7 kHz in the FER is due to the inherent notch present in the FFR (Figure 5) and must be equalized before playback. The FER has a sharp notch at around 10.6 kHz, which is labeled as the frontal perception notch (FPN) for the dummy head. This notch is of special interest, as this is attributed to the interaction of the acoustic waves with the pinna and the concha of the ear [9]. The formation of FPN is ascribed to the frontal projection of sound since this notch is inconspicuous in SER. FPN is exclusive to the individual and the position of this notch depends on the anthropometric characteristics of the ear. We can also estimate the position of the FPN from the impulse response of the FER (Figure 8). Hebrank et al. [9] had pointed out that these notches are created by the destructive interference between the direct sound and the reflected sound by the ear at the entrance of the ear canal. The delay between the first two successive peaks should correspond to the odd multiple of half time periods for destructive interference. The delay between the two successive peaks here is 340 μs corresponding to 7 half periods, which computes to 10.29 kHz as the frequency of the destructive interference. The deviation

![Graph](image-url)
from 10.6 kHz is due to the phase shift stemming from the high-pass character of the pinna reflection [19]. The energy in the largest rear boosted band (8 to 16 kHz) as observed by Blauert [1], is highest in SER compared to FER, leading to a rear perceptual bias (Figure 7). Furthermore, we observe that spectral differences between FER and SER (Figure 9) display a similar boost/attenuation pattern as Blauert’s [1] observation of the sound pressure level differences from the front and the rear directions. However they differ in the amount of gain in each of the directional bands. These findings confirm the fact that the frontal emitter naturally gives a frontal perception and the side emitter tends to give a rear perception.

5. EQUALIZATION

Headphones are not acoustically transparent as it distorts the source signal spectrum at the eardrum, and equalization is often required to remove the spectral distortion introduced by the headphone. The source is filtered with a pair of HRTFs before playing back through headphones. For a perfect 3D sound reproduction devoid of all the confusions mentioned earlier, individualized HRTFs and individualized equalization is necessary. Effective compensation is not possible even with individualized equalization because of the slight variation in the re-positioning of headphones on the listener’s head. This variation leads to a different spectrum at the eardrum, and prevents complete equalization by a pre-defined equalizer (Figure 3, 4, 5) [2]. Kulkarni et al. [18] found out that a mean inverse filter computed by averaging the response for several headphone placements performs better compared to an inverse filter computed for a given placement.

Traditionally, the frequency response of the headphone at the reproduction point (Blocked ear canal or eardrum) is inverted and is convolved with the source signal to remove the distortions introduced by the headphone (Type 1 equalization). This approach is used for the equalization of the side emitter since we intend to remove the effect of the headphone ear coupling. We introduce a different technique for the equalization of the frontal emitter as the reflections/diffractions created by the interactions with the pinna are important and should be retained even after equalization. Thus, we equalize the frontal emitter only to its FFR for playback (Figure 5) so that these individual cues are retained (Type 2 equalization).

Inversion often leads to excessive boosting of certain frequencies, which is generally unpleasant. To alleviate this problem, inverse filters were created based on Kirkeby’s fast deconvolution method using frequency dependent regularization to obtain acceptable accuracy of equalization [17].

6. ADDITIONAL PROCESSING BLOCKS

The frontal projection of the sound exaggerates the frontal spectral cues and resolves the ambiguity in the cone of confusion region. The non-individual HRTFs are partially individualized by the superposition of the auxiliary individual cues over the general HRTF cues. Additionally, decorrelation is used to externalize the virtual auditory image, and thus further enhance the frontal imagery [20]. Furthermore, front-back biasing [8] can be used to control the auditory depth of the virtual image. Figure 10 shows the complete processing block diagram.
6.1. Decorrelation

Decorrelation is a technique used to externalise the stereo sound source and produces a diffused sound field preserving the localization of the auditory image [20]. Decorrelation filters can be easily created by generating an all pass filter with a phase response constructed from combinations of random number sequences. Filtering the stereo sound source with a decorrelating filter reduces the correlation between the two channels rendering an auditory image that is more diffused and out of head.

6.2. Front Back Biasing

Blauert [1] found that certain bands (directional bands) in the frequency spectrum were boosted/attenuated, characterising the arrival direction of sound. He also noticed that these directional frequency bands had alternated boost and attenuated patterns for the sounds coming from the front and rear directions (Figure 11). This concept has been previously used in [8] to investigate the frontal perception of the auditory image in headphones. Here, we adjust the frontal bias filter to control the auditory depth of the frontal image. In the present study, we consider only the first band in the frontal bias filter, as shown in Figure 11 to illustrate the auditory depth. Several studies in the past suggest that a boost in the lower frequencies renders a closer auditory image to the listener [21]. In the proposed system, we have adopted a frontal bias filter, which spans up to 16 kHz, based on Blauert’s results for directional bands [1].

The gain in band 1 (Figure 11) in the frontal bias filter which ranges from 225 to 680 Hz can be attenuated (6 dB to 0 dB) to increase the perceived distance of the auditory image. Other bands can also be tuned in order to control the auditory depth which will be studied in detail in our future work. In the next section, we describe the subjective experiments conducted and the results supporting the theoretical validation carried out.

7. SUBJECTIVE TESTS AND RESULTS

Subjective experiments were conducted in the audio room at NTU. These tests were carried out to validate the theoretical analysis developed for the frontal projection of sound primarily with a focus on the front-back confusions and types of equalization for the frontal emitter. The contribution of the decorrelation block and the use of front-back bias filter to control the auditory depth were also investigated through these tests. A total of 15 subjects participated in the tests as volunteers, who were aged between the ages of 20-30 years. Most of the subjects had prior experience in undergoing similar tests. None of the participants had any history of hearing disorder in the past.

7.1. Subjective Test 1

The purpose of this test is to explore the difference in the rate of front back confusions in the side emitter and the frontal projection emitter. Equalization of both emitters was done before playback. The side emitter was equalized in the traditional manner (Type 1 equalization), while the frontal emitter was equalized by both Type 1 as well as with Type 2 equalization. Thus, the stimulus signal set consisted of three stimulus signals for each azimuth angle in this test. The test stimuli used were white Gaussian noise filtered with HRTFs of azimuth angles of 0°, 30°, 60° and 75°.

![Figure 11 Frontal bias filter specifications. Bands (1, 3, 5) are frontal boosted bands; Bands (2, 4) are rear boosted bands.](image-url)
7.2. Results – Subjective Test 1

Front-back confusions were mainly investigated in this test. The frontal projection helps in minimizing the front-back ambiguity by creating unique individualized spectral cues. The type of equalization also plays an important role in resolving the confusions. Figure 12 shows the percentage of front-back confusions experienced by the subjects at different azimuth angles. The responses from the subjects, who localized the sound as coming from behind, is considered to be confused since we only presented frontal sounds. It can be seen that the confusions are highest in the playback from conventional headphones (side emitters) with Type 1 equalization. Projection from the frontal emitters with Type 1 equalization reduces the rate of front-back confusions at least by 50% of those obtained from the side emitter. Using Type 2 equalization on the frontal emitter further reduces the front-back confusion to below 20% for 30°, 60°, 75° azimuth and 23.8% for 0° azimuth. Overall there is an improvement of 3 times of accuracy in 3D audio reproduction using non-individual HRTFs from the frontal emitters compared to playback from the side emitters. This also proves the effectiveness of the equalization we have employed for the frontal emitters. We note that the performance of the frontal emitter is consistent among all the frontal azimuth angles tested. Figure 13 shows the subjective score of the relative distances of the auditory image perceived for all frontal azimuths. We observe that the playback from the conventional headphone is most likely always perceived at the rear, while the frontal projection is perceived at the front of the head. It can also be seen that the proposed equalization method results in a further distant image relative to the conventional equalization as rated by the subjects.

7.3. Subjective Test 2

This test investigated the effect of decorrelation and the frontal bias filtering to control the auditory depth. Subjects listened to four different stimuli (white noise) filtered with HRTF of 30° azimuth (0° Elevation) and judged the relative position of the virtual image from the head in the horizontal plane. All the test signals were played through the frontal emitters in this test. Figure 14 depicts the subjective response chart. The different stimuli used in this test were:

1. HRTF filtered signal with Type 2 equalization (Binaural spatialization)
2. Binaural spatialization + decorrelation
3. Binaural spatialization + decorrelation + frontal bias (All bands present)
4. Binaural spatialization + decorrelation + adjusted frontal bias (Boost in band 1 attenuated to 0 dB)

Double blind test procedure was adopted in both the tests, where the evaluator as well as the subject is unaware of the order of the test signal set. In our subjective tests, we presented all the virtual sounds from the frontal azimuthal plane and thus, the subject’s responses were based on only the absolute cues provided by frontal sound with no additional cues for front-back discrimination.

7.4. Results – Subjective Test 2

7.4.1. Decorrelation block

Decorrelation has the effect of externalizing the auditory image when the source is perceived to be inside the head. Frontal projection, with our equalization has already been seen to externalize the sound source well in front of the head. Additionally, applying decorrelation has the effect of further increasing the perceived auditory distance in the front away from the head. This can be seen from Figure 15, where the relative distance from the head increases (stimulus 1 to stimulus 2) when decorrelation is applied.
7.4.2. Frontal bias filter

From the preceding experiment that showed that the decorrelation block makes the virtual auditory image more frontal, we apply the frontal bias filter to modify the auditory depth by adjusting different frequency bands (stimuli 3 and 4). From Figure 15, we note that as the frontal bias is applied (stimulus 3), the auditory image approaches closer to the head mainly due to a boost at the lower frequencies (Band 1 in Figure 11). Further it is observed that the perceived auditory distance increased away from the head as the gain from band 1 is attenuated (stimulus 4). Other bands in the frontal bias filter can be similarly modified to change the auditory depth.

8. CONCLUSIONS

In this paper, we have studied the frontal projection of sound in detail for improving 3D audio playback using non-individual HRTFs. We found that this configuration can help to reduce the rate of confusions at least by 50% irrespective of the type of equalization. We have performed detailed objective analysis of the frontal emitter and compared it with the conventional side emitter configuration. The low-frequency response of the frontal emitter has been improved compared to the response reported in [8]. A different technique for equalization of the frontal emitters was also introduced. Subjective experiments confirmed that this kind of equalization for the frontal emitter further minimizes the confusions thereby improving the accuracy of 3D audio playback system compared to the conventional headphones. Moreover, decorrelation improves the extent of frontal perception by making the auditory image more externalized. A well-controlled auditory depth perception can be achieved by carefully adjusting different frequency bands in the frontal bias filter. However, one of the limitations in this procedure is the timbre change of the original sound source. Techniques using the loudness in critical bands are currently being studied in order to obtain a trade-off between the timbral change and the accuracy of reproduction of 3D audio. In conclusion, we have found that a more convincing 3D audio effect can be reproduced through a set of non-individual HRTFs using a frontal-emitter headphone. In this new headphone structure, listener can naturally adapt the non-individual HRTFs to match to their individual HRTFs.

9. ACKNOWLEDGEMENTS

This work is supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research grant MOE2010-T2-2-040 and in collaboration with Philips.
10. REFERENCES


