



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

The role of spectral features in sound localization

Toledo, Daniela; Møller, Henrik

Published in:
Audio Engineering Society Convention Papers

Publication date:
2008

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Toledo, D., & Møller, H. (2008). The role of spectral features in sound localization. In Audio Engineering Society Convention Papers: Papers 7314-7500 Audio Engineering Society.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Audio Engineering Society Convention Paper 7450

Presented at the 124th Convention
2008 May 17–20 Amsterdam, The Netherlands

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

The Role of Spectral Features in Sound Localization

Daniela Toledo¹ and Henrik Møller¹

¹*Section of Acoustics, Aalborg University, 9220, Aalborg-Ø, Denmark*

Correspondence should be addressed to Daniela Toledo (dt@es.aau.dk)

ABSTRACT

Spectral components of head-related transfer functions (HRTFs) are highly dependent on the anthropometric characteristics of subjects. In the low frequency range, a common structure is often found in HRTFs from different subjects. However, individual differences are seen at high frequencies. In binaural synthesis with non-individual HRTFs, localization errors occur if the spectral characteristics of the directional filters used do not match the individual characteristic of the listener. This investigation is focused on the spectral characteristics of HRTFs that are relevant as localization cues and how to parameterize them. This is done by cross-matching individual and non-individual HRTFs from different subjects according to the results of localization experiments.

1. INTRODUCTION

Head-related transfer functions (HRTFs) contain the amplitude and phase information used to localize sound sources. It is known that HRTFs can be separated into temporal and spectral components, both of which are relevant in sound localization. A widely used model includes the temporal information in the interaural time difference (ITD) and represents the spectral features by a minimum-phase approximation of the HRTFs [1]. Temporal and spectral cues are direction-dependent and vary across subjects to different extents: ITDs are known

to change little and they can be approximated with different methods [2]. However, noticeable changes in spectrum are seen due to anthropometric differences [3], in particular due to differences in the shape of the pinnae.

Many applications of binaural technology use non-individual HRTFs -i.e. the HRTF filters used were not obtained from the listener. In such cases, localization performance is degraded if the spectral characteristics of the filters do not match the listener's HRTFs. The degradation can

be observed, for example, in [4] [5] [6] [7]. The degradation is seen as reversals: a source placed in one hemisphere is perceived as coming from the opposite one. Typical reversals are front-back and up-down confusions. It remains unclear which spectral features serve as localization cues and how these reversals are produced. Some literature suggest that determined frequency ranges and/or spectral features such as peaks and notches are associated to localization in specific directions [8] [9] [10] [11] [12] [13]. Characterization of head-related impulse responses (HRIRs) can also be found, for example, in [14] [15]. The hypothesis of our investigation is that particular spectral features can be identified as localization cues -either in narrow or broad frequency ranges- and parameterized by adequate descriptors. Cross-matching of individual and non-individual HRTFs is done according to the results of localization experiments, in order to identify the relevant features.

1.1. Literature review

Studies in spectral characteristics of HRTFs can be found in the literature over the past 40 years (see [16] for a thorough review). Some of these works are mentioned in the following. Care has to be taken in the comparisons, as the experimental procedures in these works differ in a great extent.

Blauert [8] analyzed localization of 1/3 octave band noise presented through five different arrangements of sound sources that evoked localization in the medium sagittal plane (MSP). Blauert divided the upper hemisphere (0° of elevation to the front of the subject to 180° of elevation to the back of the subject) into sectors front, above and behind. Since the judgements of direction clustered depending on the center frequency of the noise presented, Blauert formulated the concept of directional bands: directions in space are associated with particular frequency bands. Therefore, if the center frequency of a noise lies within these bands, subjects are more likely to perceive the sound as coming from its associated direction. Blauert also proposed the boosted bands: frequency bands with peak excitation are distinctive for sources to the front or to the back.

Hebrank & Wright [12] presented filtered noise

through a loudspeaker that moved in elevation from -30° to 210° in the MSP (with 0° to the front of the subject), in 30° steps. Filters were passband, stopband, high-pass and low-pass. They identified those frequencies for which localization judgements were biased towards a particular direction and they measured artificial pinnae -replicas made from humans- to correlate the filtering needed to the actual one provided by the pinna transfer functions (ATFs).

The findings of Blauert and Hebrank & Wright are consistent with each other. Hebrank & Wright found that perception of frontal direction is determined by a 1-octave notch between 4kHz and 8kHz with increased energy above 13 kHz. They identified that the low cut-off frequency of the notch was responsible for elevation discrimination in the frontal directions. This is in agreement with both directional and boosted bands reported by Blauert for the frontal direction -even though Blauert stated that around 8 kHz boosted and directional bands evoked behind directions. The cue for directions above was found by Hebrank & Wright to be a 1/4 octave peak between 7kHz and 9kHz. This is also in accordance to Blauert's findings of a peak centered at 8kHz as directional band for above. The cue for behind reported by Hebrank & Wright is a small peak between 10kHz and 12kHz and it is consistent with Blauert's directional and boosted bands. Blauert also found a band around 1kHz that cued the behind directions, but this frequency was not tested by Hebrank & Wright as cut-off nor center frequency. Lastly, Hebrank & Wright showed that the ATFs provided the directional filtering needed for localization. They hypothesized that the main cues were given by reflections from the back wall of the concha and formulated a model in time domain that supported their data.

Humanski & Butler [9] studied localization performance with real sources at three different azimuth angles, for elevation angles ranging from $+30^\circ$ to -30° (with 0° just to the front of the listener) in 10° steps. They focused on overt and covert features. Overt features are defined as those that can be identified in a single HRTF as its maximum or minimum. Covert features are obtained from comparison of single frequencies in the HRTFs

across directions. They analyzed the quality of covert and overt peaks and dips -i.e. whether they were uniquely contained in HRTFs from a single direction or not) and their relationship to localization of a train of high-pass filtered noise bursts with cut-off frequency at 4.5kHz. They found that, for the ipsilateral ear, covert peaks and dips and overt dips accounted well for localization. However, they suggested that covert peaks were a more robust cue since it was the only one that contributed to sound localization to the contralateral ear when the ipsilateral was occluded.

Bloom [11] reported that the perception of a certain elevation could be evoked by manipulating spectral cues. He concentrated on the notches of diffuse-field HRTFs computed from 7 measurements in the frontal plane. He filtered a 1-octave random noise signal centered at 8kHz with a notch filter with varying center frequency. A localization experiment in the frontal plane showed that, as the center frequency was increased from 6.3kHz to 10kHz, the perceived elevation went from -30° to $+45^\circ$. The contralateral ear was occluded during the experiment. He also tested elevation in the lateral angle with an arrangement in which listeners changed the center frequency of the notch when sound was coming from one of seven loudspeakers until they matched the perception of elevation they had from another loudspeaker. The resulting center frequency of the notch correlated well with the dips in the directional filters.

Middlebrooks [10] studied the localization of 1/6 octave band noise bursts signals centered at 6kHz, 8kHz, 10kHz and 12 kHz and the relationship with the spectral features of the external ear. He presented the sound from a loudspeaker that could be positioned in one of 66 locations, ranging from -160° to $+160^\circ$ in azimuth and -40° to $+80^\circ$ in elevation (with 0° to the front of the listener). Of interest to our work, he compared the directional filtering provided by the ear under two conditions: when sound was actually coming from a given direction and when sound was perceived as coming from that direction. Diffuse field HRTFs were used, and they were passed through a bank of filters that resembled the cochlea. Middlebrooks showed that the judgement of localization was determined by a

best fit to the set of directional filters: a narrow band signal exciting a determined directional filter would produce a specific signal at the eardrum, but it would be localized to a different direction. This other direction would be such that, if the corresponding directional filter was excited with a broadband signal, the result at the eardrum would be similar to that produced by the narrow band signal.

Langendijk & Bronkhorst [13] reported localization performance when spectral cues were removed from minimum-phase representations of HRTFs. Stimuli was processed by means of binaural synthesis and reproduced through individually equalized headphones. Broadband Gaussian noise was filtered with diffuse-field HRTFs from which cues in specific frequency ranges had been removed. The removal consisted on replacing specific frequency ranges by the average value within it -i.e. the spectrum was flat in the range, with only a gain factor applied-. These frequency ranges were from 2 octave to 1/2 octave bands, spanning from 4kHz to 16kHz. They showed that the frequencies that more prominently cued front-back directions were in the range from 8kHz to 16kHz -however, removing cues below 8kHz also generated front-back reversals, particularly for directions in the lower hemisphere. Frequencies that cued up-down directions were in the range from 5.7kHz to 11.3kHz. Other observations made by Langendijk & Bronkhorst were that frontal directions contained a peak in the range from 8kHz to 16kHz that was not present in rear directions. The idea that elevation was cued by a notch in the range 5.7kHz to 11.3kHz (with increasing center frequency as elevation increases) was not supported by the experimental results, since removing that frequency band did not affect localization. According to their results, the authors also suggested that cues below 4kHz were not as relevant in localization as cues above that frequency.

2. METHODS

The objective of the experiments is to obtain the evoked localization of sound sources for a given set of individual and non-individual HRTFs, for a small number of subjects. This means that each pair of non-individual HRTFs corresponds to a direction for which it was measured -original

direction-, but for each subject it is associated to an evoked direction which may be other than the original one. Localization performance results are used to compare individual HRTFs from a given original direction with those non-individual HRTFs that evoke the same direction. This is the basis of the cross-matching that will allow finding relevant spectral cues.

The listening experiments test localization performance for three conditions: real life, binaural synthesis with individual HRTFs and binaural synthesis with non-individual HRTFs. The testing environment consists on the same anechoic chamber used for measuring HRTFs, and it is unchanged across conditions. The investigation is constrained to the MSP as the interaural differences are minimized. Real life condition is included for two reasons. Firstly, it has been shown elsewhere that localization performance in binaural synthesis with individual HRTFs is close to localization in a real life condition [17]. We check the validity of this assumption for our subjects. Secondly, we are interested in how good localizers the participants are.

2.1. Coordinate system

Azimuth corresponds to lateral angles and elevation to vertical angles. In the coordinate system chosen, 90° and -90° azimuth are situated at left and right sides, respectively. All directions are given in (azimuth ϕ , elevation θ). $(0^\circ, 0^\circ)$ is to the front and $(180^\circ, 0^\circ)$ is to the back of the subjects.

2.2. Subjects

Subjects with normal hearing participate in the experiment. Their hearing threshold are determined by a standard pure-tone audiometry in the frequency range from 250 Hz to 8 kHz. None of the subjects have hearing thresholds above 15dB HL. The total number of subjects participating in the experiment has not been defined at the time of submission.

2.3. Loudspeaker setup

HRTFs measurements and listening experiments are conducted in an anechoic chamber. Fifteen loudspeakers Vifa M10MD-39 mounted in spherical cabinets with diameter 15.5cm are used in the setup. A typical frequency response of one of this loudspeakers can be found in [3]. Loudspeakers are positioned in a vertical arc at 0° azimuth, with a separation of

22.5° in elevation. The distance from the center of the head of the subjects to the loudspeakers ensures far field propagation.

2.4. Signal generation and control

The equipment is placed in a control room next to the anechoic chamber. Signals are played back through a PC with a digital sound card RME HDSP 9632 connected to an external AD/DA converter RME ADI-8 DS. For HRTFs measurements and listening tests in real life condition, the signals feed a power amplifier Pioneer A-616 modified to provide 0dB gain. The output of the amplifier is connected to a custom made switch, controlled through the PC, which feeds the signal to the corresponding loudspeaker. For listening tests in virtual condition, the signals from the AD/DA converter are routed to a headphone amplifier Behringer Powerplay Pro HA 4400 and played back through headphones Beyerdynamic DT990, individually equalized. Typical transfer functions for this headphone are shown in [18].

2.5. HRTFs measurement

The general description of the procedure to measure HRTFs can be found in [3]. In this investigation, the method is applied with small differences. Briefly, a dual channel MLS system developed at our laboratory [19] is used. The signals are collected by two Sennheiser KE 4-211-2 miniature microphones placed at the blocked entrance of the ear canals of the subjects. Microphones are calibrated and connected to a power supply that provides a gain of 20dB. Two measuring amplifier Brüel & Kjær 2607 are used before feeding the signals to the AD/DA converter and back to the PC. Free-field HRTFs require a reference measurement for the pressure division. This is done at the position of the center of the head, with the subjects absent. All measurements are done at a sampling frequency of 48kHz and appropriate post-processing is implemented with MATLAB.

2.6. Stimuli

The stimuli used for the listening experiment is broadband noise of 1s length.

For testing in real life condition, the stimuli is equalized to account for the frequency response of the loudspeakers. The transfer characteristic of each loudspeaker is measured in order to build

minimum-phase inverse filters. These are applied to the signals being played back. All filtering is done offline.

In the virtual sources condition, the stimuli is filtered with the appropriate HRTFs -individual or non-individual- and processed with individual filters to equalize the headphones response.

2.7. Selection of representative non-individual HRTFs

Non-individual HRTFs to be tested are selected from a large database of measured HRTFs [3]. HRTFs measured from subjects participating in the experiment are also considered.

2.8. Procedure

Real life and binaural synthesis are tested in separate groups. Presentations within a group are fully randomized and divided into blocks. The binaural synthesis group comprises both individual and non-individual HRTFs. Directions tested are selected from the available loudspeakers in the vertical arc, but they have not been defined at the time of submission. Evoked localization judgements are obtained through an absolute localization paradigm. Subjects are presented with a 2D graphical interface in which they have to reflect the perceived localization in azimuth and elevation.

3. RESULTS

The experiment is currently in preparation. Results are presented in the Convention.

4. REFERENCES

- [1] J. Plogsties, S. Krarup Olesen, P. Minnaar, F. Christensen, and H. Moller, "Audibility of all-pass components in head-related transfer functions," in *Proceedings of 108th Audio Engineering Society Convention*, no. 5132, (Paris), pp. 1–14, Audio Engineering Society, February 19-22 2000.
- [2] P. Minnaar, J. Plogsties, S. Krarup Olesen, F. Christensen, and H. Moller, "The interaural time difference in binaural synthesis," in *Proceedings of 108th Audio Engineering Society Convention*, (Paris, France), Audio Engineering Society, February 19-22 2000.
- [3] H. Moller, M. Friis Sorensen, D. Hammershoi, and C. Boje Jensen, "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–321, 1995.
- [4] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using non-individualized head-related transfer-functions," *Journal of the Acoustical Society of America*, vol. 94, pp. 111–123, July 1993.
- [5] H. Moller, D. Hammershoi, C. Boje Jensen, and M. Friis Sorensen, "Evaluation of artificial heads in listening tests," *Journal of the Audio Engineering Society*, vol. 47, no. 3, pp. 83–100, 1999.
- [6] H. Moller, M. Friis Sorensen, C. Boje Jensen, and D. Hammershoi, "Binaural technique: Do we need individual recordings?," *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.
- [7] P. Minnaar, S. K. Olesen, F. Christensen, and H. Moller, "Localization with binaural recordings from artificial and human heads," *Journal of the Audio Engineering Society*, vol. 49, pp. 323–336, May 2001.
- [8] J. Blauert, "Sound localization in median plane," *Acustica*, vol. 22, no. 4, pp. 205–&, 1969.
- [9] R. A. Humanski and R. A. Butler, "The contribution of the near and far ear toward localization of sound in the sagittal plane," *Journal of the Acoustical Society of America*, vol. 83, pp. 2300–2310, June 1988.
- [10] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *Journal of the Acoustical Society of America*, vol. 92, pp. 2607–2624, Nov. 1992.
- [11] P. J. Bloom, "Creating source elevation illusions by spectral manipulation," *Journal of the Audio Engineering Society*, vol. 25, no. 9, pp. 560–565, 1977.

- [12] J. Hebrank and D. Wright, "Spectral cues used in localization of sound sources on median plane," *Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.
- [13] E. H. A. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002.
- [14] H. L. Han, "Measuring a dummy head in search of pinna cues," *Journal of the Audio Engineering Society*, vol. 42, no. 1-2, pp. 15–37, 1994.
- [15] Y. Hiranaka and H. Yamasaki, "Envelope representations of pinna impulse responses relating to 3-dimensional localization of sound sources," *Journal of the Acoustical Society of America*, vol. 73, no. 1, pp. 291–296, 1983.
- [16] S. Carlile, R. Martin, and K. McAnally, "Spectral information in sound localization," *Auditory Spectral Processing*, vol. 70, pp. 399–434, 2005.
- [17] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening .2. psychophysical validation," *Journal of the Acoustical Society of America*, vol. 85, pp. 868–878, Feb. 1989.
- [18] H. Moller, D. Hammershoi, C. B. Jensen, and M. F. Sorensen, "Transfer characteristics of headphones measured on human ears," *Journal of the Audio Engineering Society*, vol. 43, pp. 203–217, Apr. 1995.
- [19] S. Olesen, J. Plogsties, P. Minnaar, F. Christensen, and H. Moller, "An improved mls measurement system for acquiring room impulse responses," in *Proceedings of NOR SIG 2000*, (Kolmden, Sweden), pp. 117–120, IEEE Nordic Signal Processing Symposium, 2000.