Detection of Notches in Head-Related Transfer Functions

R. BOMHARDT[1]

[1] Institute of Technical Acoustics, Medical Acoustics Group, RWTH Aachen University, Germany

rbo@akustik.rwth-aachen.de

Abstract. *Binaural sound reproduction is becoming more and more popular wherefore it is worthy to provide this technology for a broad audience. The usage of head-related transfer functions of artificial heads is not optimal because the interaural differences and the monaural cues of these transfer functions do not fit with those of an individual listener. Due to this mismatch often in-head localization or front-back confusions occur. To identify mismatches between a set of head-related transfer functions of an artificial head and a subject, this paper presents an algorithm that detects notches, which are relevant for the localization of elevated sources, by the use of a kalman filter. Meanwhile the frequency independent maxima of the transfer functions result from the resonances in the cavum conchae, the notches result from destructive interferences. For this reason the notches are frequency and elevation angle dependent so that the auditory system can use them to localize sources for instance in the median plane.*

Keywords

Binaural Hearing, Head-Related Transfer Function, Spatial Audio.

1. Introduction

Due to the fact that binaural reproductions[1] with artificial heads often result in in-head localizations or front back confusions, this paper presents with an algorithm which analyses a set of head-related transfer functions (HRTFs)[2] with respect to its resonances and destructive interferences [1]. Therefore, the presented algorithm ensures a comparison between two different HRTF datasets.

In general, humans use monaural cues, which are characterized by the resonances and destructive interferences for the source localization, particularly on the cones of confusions. Such a cone of confusion is for instance the median plane where the interaural level and time differences are zero and

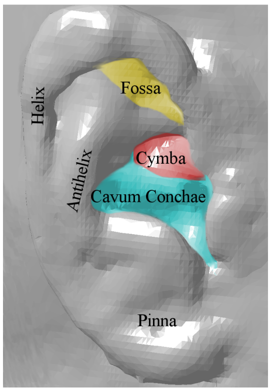

Fig. 1. The pinna of a human ear is scanned by MRI and reconstructed. The cavum conchae is the biggest deepening in the ear which merges into the ear canal. Additionally, the cymba and fossa, which influence the resonances at higher frequencies, are displayed. The notches are mainly affected by the illustrated rim of the helix and antihelix.

are not beneficial for the auditory system. On all other cones, there the level and time differences are equal and cannot be used for the evaluation of the source direction neither.

The resonances of the cavum conchae, which is shown by an image of the pinna in figure 1, can be observed frequency dependent above 5 kHz [2][3]. For most of the human ears, the standing wave of the first resonance, despite the ear canal resonance at 3 kHz, has a maximum in the complete deepening of the cavum conchae. The second and third one are roughly split by the cymba and can be found around 10 kHz. Above, the shape of the standing wave is more complex and is additionally influenced by the fossa. These resonances produce a colouration of perceived sound which is especially noticeable when a non-individual HRTF dataset is used.

Besides these resonance phenomenons, which are not dependent on the angle of the incident wave, the notches, which are caused by destructive interference in the cavum conchae, can be observed as angle dependent. These frequency and elevation dependent notches enable humans to localize sources on the cone of confusion. For most of the humans two to three notches in the range of the human hearing can be observed. Mostly, the first one is located between 5 and 12 kHz (see figure 2) and the second one starts around 10 kHz for an elevation angle of $-50°$. Satarzadeh et Al. [4] assume that these notches result from an incident wave which either reflects directly on the rim of the cavum conchae (antihelix) or passes the rim and reflects on the helix indirectly. From

[1] Binaural is related to hearing with two ears. For a binaural reproduction often headphones are used where a binaural recording, consisting of a recording at left and right ear, can be played back channel separated.

[2] These transfer functions are measured dependent on the direction in a constant distance on a sphere with a loudspeaker array in respect to a centred reference position.

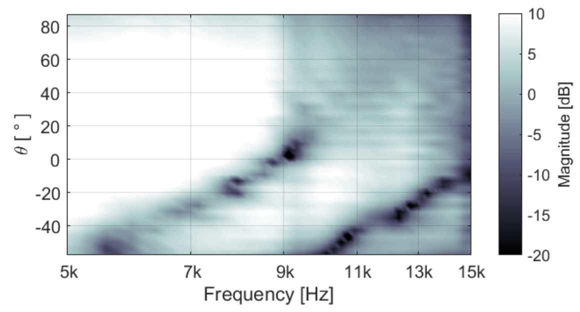

Fig. 2. A set of HRTFs of a subject is shown at azimuth $\varphi = 300°$ for the elevation angles θ in the frequency domain. The colour displays the magnitude level in dB.

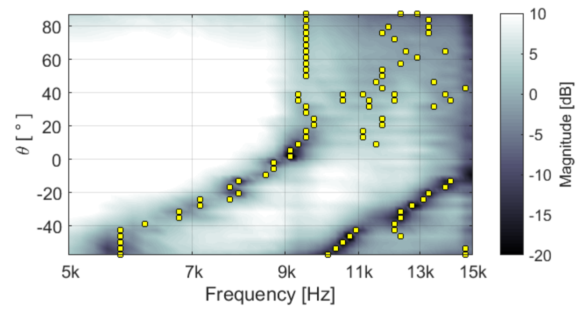

Fig. 3. The local minima in a spherical HRTF slice for the horizontal direction $\varphi = 300°$ are shown in the image as yellow dots.

the pinna image in figure 1 can be derived that the shortest path from the antihelix and helix to the ear canal entrance for an incident wave is dependent on the elevation angle and therefore also frequency dependent. This enables humans to localize elevated sources.

Due to the importance of the notches, Spagnol et Al. [6] used a tracking algorithm to detect the notches in the median plane. Afterwards they tried to identify the corresponding contours of the ear by image processing, from which they conclude that especially the rim of the cavum conchae, the antihelix and the helix play a major role for the localization on the cones of confusions. In addition, Raykar et Al. [7] came to the same conclusion that the antihelix and the rim of the helix are important for the notches. In contrast to Spagnol et Al. [6], they used a minimum detection by neglecting the spectrum above 0 dB.

Since these notches are not only relevant in the median plane but also on all other cones of confusion, this paper presents an approach that detects and compares these notches for different azimuth angles. Similar to the used motion tracking technique of Spagnol et Al. [6], here the kalman filter is used to detect the linear curve progression of the notches on a logarithmically scaled axis (see figure 2). However, the observation of the notches is disturbed by measurement noise and the resonances of the cavum conchae wherefore the algorithm is able to reject frequency independent notches. For instance, such a notch can be observed in figure 2 from 20° to 60° at 10 kHz[3]. Afterwards the position of the frequency and elevation angle dependent notches can be linearised. Finally, the notches are evaluated and analysed dependent on the elevation and azimuth angle.

2. Estimation of Notches

The HRTF datasets of twelve subjects and an artificial head are measured from $\theta = -58°$ to $90°$ with a loudspeaker array on a turntable (see Bomhardt and Fels [5] for details). These datasets are used for the analysis of the maxima and notches. For the individual dataset in figure 2 can be seen that there are two notches: One starts at $(\theta, f) = (-50°, 6\,\mathrm{kHz})$ and disappears approximately at $(20°, 10\,\mathrm{kHz})$ and the other one starts at $(-50°, 10\,\mathrm{kHz})$ and disappears at approximately $(0°, 15\,\mathrm{kHz})$. This is very characteristic for most of the measured individual HRTF datasets and will be used for the initialization of the kalman filter. Additionally, it is observed that the magnitude of the notches is approximately $-10\,\mathrm{dB}$ whereas the one of the maxima is larger than $10\,\mathrm{dB}$. Especially for angles $\theta > 0°$, often a frequency independent minimum at $8\,\mathrm{kHz}$ can be observed which results from the two resonances at 6 and $9\,\mathrm{kHz}$ and not from destructive interference.

2.1. Detection of Minima

The local minima are detected by a maximum detection. The magnitudes of the transfer functions are lowered by $20\,\mathrm{dB}$, to avoid zero divisions, and inverted. Following, the local maxima whose magnitudes are greater than $-\frac{1}{15}\,{}^{1}\!/\mathrm{dB}$ and their drop is at least $10^{-5}\,{}^{1}\!/\mathrm{dB}$ are detected. Finally, the frequency of the notch and direction θ of the current transfer function is stored so that the notches can be represented as shown in figure 3.

2.2. Estimation of Notches by a Kalman Filter

In this paper, the kalman filter with a constant velocity approach is chosen to find the related notches in a set of HRTFs [8].

For the analysis, the magnitude of the spectra of a spherical slice for a horizontal direction φ is logarithmically transformed due to the logarithmical human sound perception. Consequently, the relationship between the frequency positions and the elevation angle of the notch is approximately linear.

The idea is that an initial position x_0 is chosen where the notch can be easily detected. In all measured HRTF datasets, the starting point can be determined around either $\theta = -50°$ for frequency independent or $\theta = 90°$ for the destructive

[3] In this paper the 0° elevation angle is defined in front of the head and positively to the top of the head.

minima. Then, the initial states x_0 for the angle are defined as:

$$x_0 = [\theta_0 \quad f_0]^T . \tag{1}$$

Based on this position, the kalman filter will estimate the following positions of the notch for higher elevation angles θ which are sometimes disturbed by resonances or measurement noise. Considering that the position of the frequency of the notch f_k will rise with a rising elevation angle θ_k, it can be assumed that the next notch position x_{k+1} in relation to the previous one x_k can be found at $x_{k+1} = x_k + \dot{x}_k \cdot \tau$. Consequently, the resulting state transition model is defined as:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{2}$$

with $\tau = 1$, so that the estimation of the next position x_{k+1} can be calculated as follows:

$$x_{k+1} = A \cdot x_k + w_k. \tag{3}$$

Additionally, a process noise w_k, which is derived from the covariance matrix $Q_{noise} = \text{diag}\,(\theta_{motion}, f_{motion})$, is considered in this equation to cover the uncertainties of the estimation.
In general the estimation process of the kalman filter for the notch detection consists of three steps:

1. Prediction of the next measurement point,
2. Nearest neighbour search for the closest point in relation to the predicted point,
3. Correction of the model by the closest point.

With the estimated position x_{k+1} of equation 3 and a measured position z_k, which can be found by a nearest neighbour search from the detected minima, the model can be corrected by the measurement equation:

$$z_k = H \cdot x_k + v_k. \tag{4}$$

Here, it is assumed that the observed position z_k is affected by measurement noise v_k which can be derived from the measurement noise covariance matrix $R_{noise} = \text{diag}\,(\theta_{meas}, f_{meas})$. The measurement model H takes only the estimated position into account wherefore it is defined as $H = [1 \quad 0]$.
The final step, where the kalman filter is updated, uses a gain factor K which is calculated from error matrices for the initial error $P_{error} = \text{diag}\,(\theta_{est}, f_{est})$, the measurement model H and measurement noise covariance matrix R_{noise} [8]. By the help of this gain factor K, the updated state estimate $x_{k|k}$ and estimated covariance matrix $P_{k|k}$ is determined for the estimation of the next position.

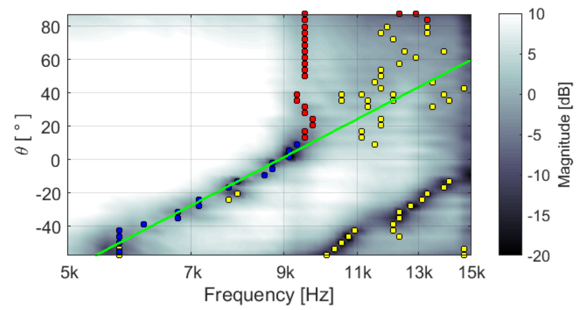

Fig. 4. The notches which are found by the local minimum search (yellow dots), the detected frequency independent notches (red dots) as well as the notches which are created by destructive interference (blue dots) are shown for a spherical slice at $\varphi = 300°$. Additionally, the green line is the approximated notch position.

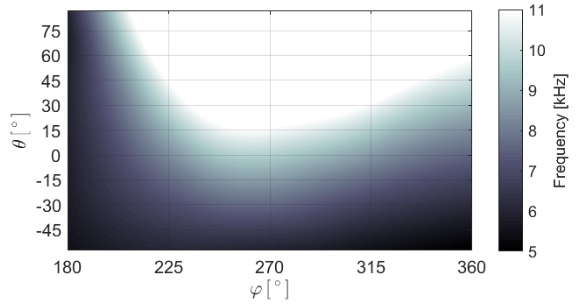

Fig. 5. The frequency of the notch is displayed as the colour over the azimuth φ and elevation angle θ.

3. Evaluation of Notches

As previously discussed, before the destructive notches can be detected, the ones which are caused by the resonances should be eliminated. Therefore, the kalman filter is also used to detect them from the largest elevation angle $\theta \approx 90°$ down to $10°$. These limiting angles are empirically determined by the analysis of the 13 HRTF datasets. The search direction, from large to small elevation angles, is chosen by the reason that the initial position of the frequency independent notches can determined more precisely at $\theta = 90°$ than at $\theta = -50°$.
The elimination of this points enables the observation of the frequency dependent notch in figure 4 more accurately. For the observation, the initial position for the kalman filter is defined at the lowest elevation angle $\theta \approx -50°$ and is limited at $\theta = 50°$ as long as additional positions can be detected. In case that the frequency independent minima could not be eliminated, the linear approximation of the destructive notch will be worse than in figure 4.
As previously mentioned, most of the studies discuss these notches in the median plane (c.f. [6]). With the current algorithm, it is also possible to investigate the destructive notches dependent on the azimuth and elevation angle. Based on the results in figure 5, it can be observed that the destructive notches are in a range of $\varphi = 200°$ to $360°$. Below $\varphi < 200°$ the notches are concealed by resonances or background noise. In figure 5 is shown that in general the fre-

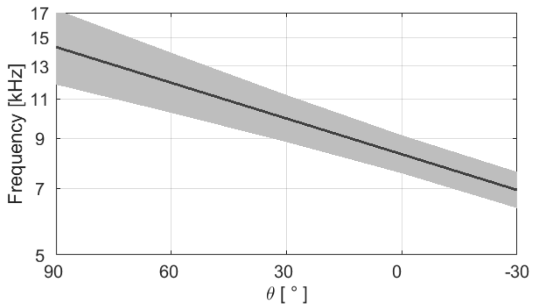

Fig. 6. The estimated and linearised position of the first notch of the 13 HRTF datasets at $\varphi = 300°$ is shown in dependency of the elevation angle θ and the frequency. The dark line is the mean line and the gray area marks the standard deviation of all linearised notch positions.

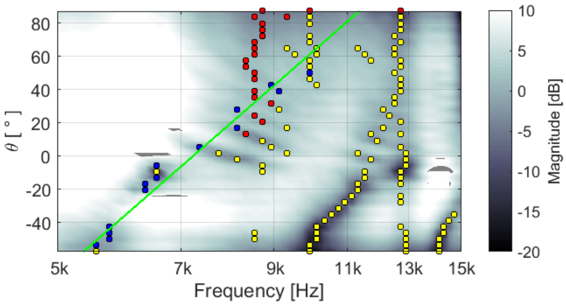

Fig. 7. The image shows the described algorithm in a case where the disturbance of the resonance maxima and noise is too high wherefore the frequency dependent notch cannot be estimated very precisely.

quency of the notch will increase for a rising elevation. Furthermore, this relationship can be observed for all azimuth angles $\varphi > 250°$. Additionally, the frequency of the notch will decrease for an increasing horizontal angle $\varphi > 250°$ and a constant elevation angle θ.

In the comparison of different individual HRTF datasets in figure 6, it can be observed that the estimated and linearised position of notch can fluctuate between different subjects. The constantly increasing frequency of the notch with a rising elevation angle shows a standard deviation of $1 - 2\,\text{kHz}$ dependent on the frequency. If such a deviation is present, either the elevation angle is incorrectly perceived or it results in a front-back confusion.

4. Conclusion

For the systematic observation of the notches, which are often disturbed by resonance maxima or noise, the presented minimum search in combination with a kalman filter shows good results. Because in figures 3 - 5 only one individual HRTF datasets is shown, it should be noted that the algorithm was tested for all other datasets as well. The algorithm fails only for two of 13 HRTF datasets to determine the notches accurately due to the disturbing resonance maxima (see figure 7).

Although the measurements are more disturbed by resonances and noise at higher frequencies, for some of the HRTF datasets it is also possible to observe the second notch around 12 kHz with the presented algorithm.

Furthermore, the differences between two HRTF datasets and their mismatch can be pointed out by the position of the notches. For a given HRTF database[4], the analysis of the position of the notches can be a criterion to choose an HRTF dataset.

Additionally, the described algorithm also allows conclusions on the colouration by the resonances of the pinna. As shown for the resulting minima due to the resonance maxima, the algorithm can also be applied on the magnitude of an HRTF dataset to find the frequency independent resonance maxima.

Finally, with the estimation of the notches, it is also possible to study the dependency between them and the dimensions of the pinna. As a consequence a better understanding, which factors play a major role for the front-back confusions, is derivable [9][10].

Acknowledgements

The project on which this paper is based was funded by the German Federal Ministry of Education and Research (BMBF), funding code 01IS14027B. Responsibility for the contents of this publication rests with the author.

References

[1] Blauert, Jens (1974): Räumliches Hören, Hirzel: Stuttgart.

[2] Shaw, Edgar A. G. (1997): Acoustical features of the human external ear. In: Binaural and spatial hearing in real and virtual environments 25, S. 47.

[3] Takemoto, Hironori; Mokhtari, Parham; Kato, Hiroaki; Nishimura, Ryouichi; Iida, Kazuhiro (2012): Mechanism for generating peaks and notches of head-related transfer functions in the median plane. In: The Journal of the Acoustical Society of America 132 (6), S. 3832-3841.

[4] Satarzadeh, P.; Algazi, V. R.; Duda, R. O. (2007): Physical and filter pinna models based on anthropometry. Audio Engineering Society Convention 122: Audio Engineering Society.

[5] Bomhardt, R.; Fels, J. (2014): Analytical interaural time difference model for the individualization of arbitrary Head-Related Impulse Responses. Audio Engineering Society Convention 137. USA, Los Angeles, 9-12.10.2014. Audio Engineering Society. 137. Aufl. USA, New York: Audio Engineering Society.

[6] Spagnol, Simone; Geronazzo, Michele; Avanzini, Federico (2013): On the relation between pinna reflection patterns and head-related transfer function features. In: Audio, Speech, and Language Processing, IEEE Transactions on 21 (3), S. 508-519.

[7] Raykar, Vikas C.; Duraiswami, Ramani; Yegnanarayana, B. (2005): Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. In: The Journal of the Acoustical Society of America 118, S. 364-374.

[8] Welch, Greg; Bishop, Gary (2006): An introduction to the kalman filter. Department of Computer Science, University of North Carolina: Chapel Hill, NC, unpublished manuscript.

[4] To avoid mismatches sometimes HRTF databases are provided to choose the most appropriate HRTF dataset for an individual listener.

[9] Zhang, M.; Kennedy, R. A.; Abhayapala, T. D.; Zhang, W. (2011): Statistical method to identify key anthropometric parameters in HRTF individualization. Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on: IEEE.

[10] Iida, Kazuhiro; Ishii, Yohji; Nishioka, Shinsuke (2014): Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae). In: The Journal of the Acoustical Society of America 136 (1), S. 317-333.

About Authors...

Ramona Bomhardt is research assistant at the Institute of Technical Acoustics at RWTH Aachen University, Germany. She received her Diploma in electrical engineering in 2010 and then worked one year in the automotive industry. In 2011 she returned to the Institute of Technical Acoustics and works in the field of binaural acoustics.