Applied Acoustics 71 (2010) 1211–1216

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical Note

Estimation of the direction of auditory events in the median plane

Elena Blanco-Martin *, Francisco Javier Casajus-Quiros, Juan Jose Gomez-Alfageme, Luis Ignacio Ortiz-Berenguer

Universidad Politecnica de Madrid, Audiovisual and Communications Engineering, E.U.I.T. Telecommunication, Crta. Valencia, km 7, 28031 Madrid, Spain

ARTICLE INFO

Article history:
Received 16 December 2008
Received in revised form 12 July 2010
Accepted 13 July 2010
Available online 12 August 2010

Keywords:
Psychoacoustics
Localization
Median plane
Spectral cues
Sound event

ABSTRACT

An instrumental method based on parametrically described spectral cues is proposed to estimate the direction of auditory events in the median plane. In contrast to a previous method where the shapes of two spectra are compared, the current method is based on known psychoacoustic features of the auditory system. The algorithm is described and the results of verification experiments are reported.

1. Introduction

Listeners are able to locate distinct auditory events in sound fields reproduced by one or more loudspeakers. The direction of the auditory events may or may not be the same as the direction of the loudspeakers employed. Therefore, there is a demand for developing instrumental measures to estimate the direction of the auditory events that combine the best aspects of the perceptual validity of listening tests and efficiency and the repeatability of instrumental measurement. Such modeling approaches, usually based on psychoacoustic knowledge, are about to become valuable tools for audio research.

The localization of auditory events depends on a combination of several parameters, among others, the nature of the sound sources, the anthropometry and hearing characteristics of the listener, voluntary or involuntary motion of sources or listeners, and the acoustic environment surrounding the listeners. Whereas some of these parameters improve localization, others, such as echoes and reverberation are often detrimental. The work described here starts with a reference environment that consists of non-personalized head-related transfer functions (HRTFs) and sound sources in a free field, that is, with no echoes or reverberation.

The simulation system creates the ear-input signals as received by a dummy head from a loudspeaker placed 1.5 m away from the dummy in the median plane. The dummy head is fixed in space and essentially symmetric, that is, the left and right ear-input signals are considered to be identical. This situation (diotic listening) is known to lead to auditory events in the median plane. It is the task of the estimation system to predict of the angle of elevation, φ, of the auditory events.

The estimation system uses two HRTF databases for comparison. One of them is the well-known KEMAR database [1], and the other one is a laboratory-built one. The latter one was built by measuring the HRTFs of HATS, a dummy head produced by Head Acoustics GmbH. Measurement was performed with a cross-spectrum method, making use of the B&K PULSE system [2]. In the following, the estimation system is described in more detail and results of predictions for sound sources at different elevations are reported. The influence of the test signals on the results is discussed by comparing white noise and speech results.

2. Sound localization in the median plane

As is known from literature the auditory system analyzes spectral cues of the ear-input signals to form the direction of auditory events in the median plane. Previous studies [3–8] show that linear distortions as caused by the pinnae, predominantly in the high frequency range above 5 kHz, act as prominent cues for median plane localization. Fig. 1 shows that the spectrum of the ear signals changes systematically in the frequency range above 5 kHz when the elevations of the sound sources is varied.

In general, spectral cues used for localization mainly appear between 4 and 16 kHz. The cues that determine the forward direction consist of a one-octave notch with a lower cut-off frequency between 4 and 8 kHz (defined as N1 by Iida [5]) and an increase in energy above 13 kHz. The overhead cue is a ¼-octave peak

* Corresponding author. Tel.: +34 91 336 55 05; fax: +34 91 336 77 84.
E-mail address: eblanco@diac.upm.es (E. Blanco-Martin).

0003-682X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2010.07.004

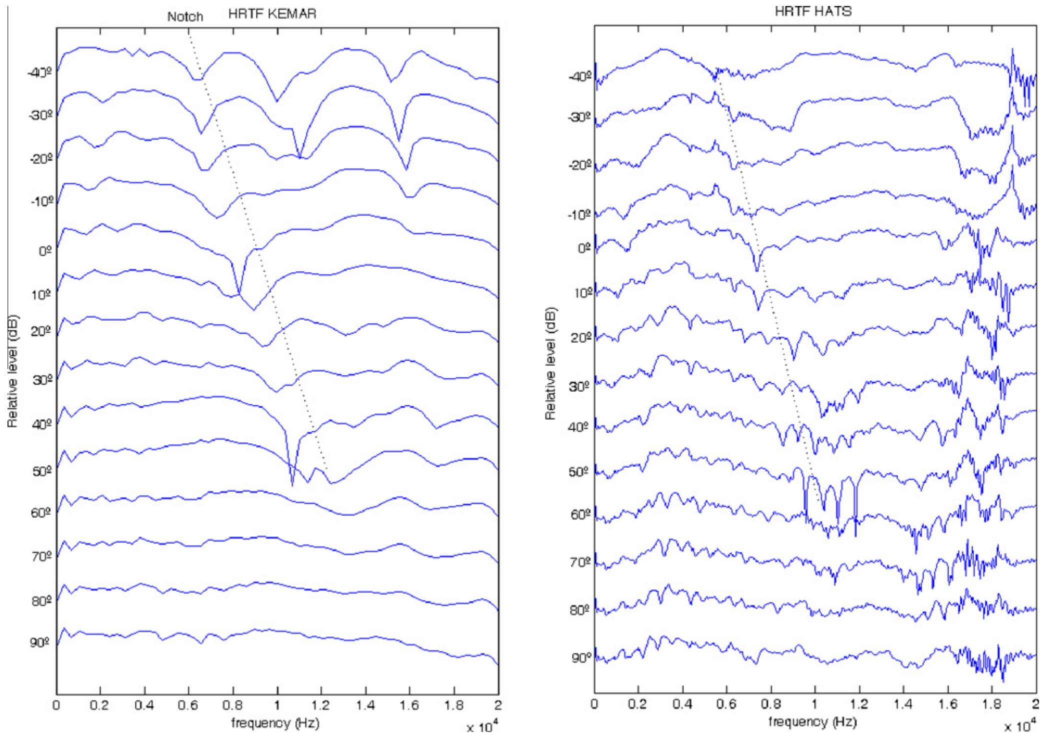

Fig. 1. Module of HRTF from the KEMAR and HATS dummy heads for an elevation angle range of −40° to 90°.

between 7 and 9 kHz (defined as P1 by Iida [5]). Note that listeners can well detect the following changes in the spectral cues: a spectral notch that changes from 6 to 10 kHz as the elevation of a sound source changes from −45 to 45° and the positions of spectral peaks and notches. The spectral positions of the peaks and notches can, of course, be tracked back to pinna-geometry.

Iida [5] has analyzed peaks and notches in the head-related transfer functions (HRTFs) with respect to the question of which of them play relevant roles as localization cues. To this end, listening tests were carried out with sound signals being generated by the use of parametrically simulated HRTFs. In those HRTFs, notches and peaks were successively introduced to examine their respective importance. Results showed that parametric HRTFs composed of the first and second notches (N1 and N2) and the first peak (P1) provided almost the same localization as natural HRTFs. Regarding the spectral peaks and notches, the spectral positions of N1 and N2 vary substantially with the elevation of the sound sources, whereas P1 did not depend on source elevation. Consequently, the first two notches are assumed to be relevant spectral cues for the formation of the elevation of auditory events. Based on this assumption, a method has been constructed for estimating the elevation of auditory events from the ear-input signals of a dummy head when presented with sound sources in the median.

High-frequency sounds give rise to auditory events at higher elevation angles [9]. This fact is related to the relatively large energy between 7 and 9 kHz that appears for "above angles" due to the respective HRTF characteristics. Use of this knowledge is made by through a database of spectral cues in our estimation system. In detail, the spectral cues from incoming sounds are compared with the spectral cues in the HRTF database. Those angles from the database that possess similar spectral cues (the "nearest" set of spectral cues) are taken as estimates for the elevation angles (Fig. 2).

2.1. HRTF databases

HRTFs were measured in 10° steps between −40° and 90° in the median plane in an anechoic chamber. The KEMAR database [1,10], having a sampling frequency of 44,100 Hz and a spectral resolution of 65 samples, was used. Also used was the HATS database [2]. Here, the sampling frequency was 65,536 Hz with a spectral resolution of 1024 samples.

2.2. Extraction of spectral peaks and notches

As mentioned above, spectral peaks and notches in frequencies above 5 kHz are assumed to contribute most prominently to the

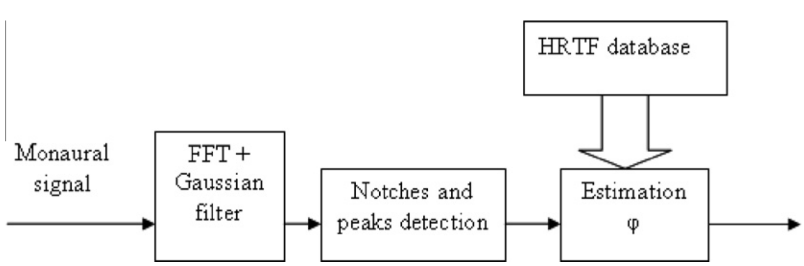

Fig. 2. Sound localization method in the median plane.