MEASUREMENT OF HEAD-RELATED TRANSFER FUNCTIONS BASED ON THE EMPIRICAL TRANSFER FUNCTION ESTIMATE

Elena Grassi, Jiwanjot Tulsi and Shihab Shamma

Institute for Systems Research, Dept. of Electrical and Computer Engr.
University of Maryland, College Park, MD 20742, USA
`egrassi@umd.edu, sas@eng.umd.edu`

ABSTRACT

An experimental procedure and signal processing method to measure Head Related Transfer Functions (HRTFs) is reviewed. The technique based on Fourier analysis system identification has an advantage over the commonly used maximum-length sequence and Golay methods if nonlinear distortions are present in the loudspeakers and their power amplification circuits. The method has been used to produce a new public HRTF database. Compared to existent public domain databases, these transfer functions have been measured at points spaced more densely and uniformly around the subject, which poses an advantage for fitting and interpolating methods. The measured HRTFs (seven subjects to date) are available by request from the authors.

1. INTRODUCTION

Our ability to localize sounds relies on several cues that are extracted from the sounds we hear [1]. Among these cues are spectrum, intensity and time differences between sounds arriving at the ears. The incoming sounds are transformed in ways which depend on the shape and size of our head, torso, and in particular of our ears. These anatomical features vary greatly across individuals and make the acoustical transformations, also known as HRTFs, highly personalized.

HRTFs can be measured and synthesized in the form of linear time invariant filters and their use is important for rendering realistic virtual audio and auditory displays through headphones. Measuring HRTFs is an expensive, time consuming endeavor and this has motivated studies of localization performance using non-individual HRTFs. Localization, particularly in the vertical dimension, is degraded when using non-individualized HRTFs [2]. Some efforts to customize general HRTFs to a particular subject have been made in [3, 4]. Reproduction of room reverberation also appears to enhance localization [5] but it seems that individually measured HRTFs are still unmatched in localization accuracy.

System identification techniques that have become popular in synthesizing HRTFs are the so called maximum length sequence (MLS) and Golay code methods. MLS is well described in [6], where is presented as a tool to obtain room acoustic transfer functions. In general, system identification requires that an input signal (excitation) be applied to the system under scrutiny while the output is being measured. From there, the transfer function of the system, i.e., the way the output relates to the input, is derived. MLS/Golay methods borrow their names from the input signals they use (the MLS/Golay codes) which are binary sequences with certain correlation properties: MLS signals have the property that their circular autocorrelation is an impulse (except for a negligible offset); Golay codes are presented in pairs of complementary sequences, with the property that the sum of their autocorrelation is an impulse. By exploiting this property the impulse response (a time domain version of the transfer function) of a **linear** system can be recovered by simple cross-correlation of the output signal with the input MLS/Golay code.

Despite their mathematical elegance, these methods may not be the most appropriate ones for measuring acoustical transfer functions in certain situations. For example, [7] describes sensitivity of Golay methods to small time variations (subject's movement) in the measurement of long impulse responses, like the ones needed to capture room acoustics; [8] analyzes how loudspeaker nonlinearities (e.g. rate limit and second/third order memoryless nonlinearity) can produce spurious peaks in the estimated transfer function. MLS signals are also known in the system identification literature as pseudo-random binary signals or PRBS [9] and to avoid confusion between input signals and the method used in obtaining the HRTFs we use the name PRBS from here on when referring to signals of this kind. It should be emphasized that some of the potential problems with MLS/Golay codes do not arise from the use of PRBS signals themselves but rather from the way the HRTFs are extracted.

The Empirical Transfer Function Estimate (ETFE) [9] used here for HRTF computation is simply based on Fourier Transform ratios. The advantage of the method lies in that it can be applied with a broad range of input signals (not restricted to PRBS) thus allowing to consider nonlinear distortion of loudspeakers and their power amplification circuits as part of the input signal rather than the system. Methods based on Fourier transform ratios have been applied before in the context of HRTFs. Examples of their use include [10], where repetitive impulses are used as excitation signal to measure HRTFs, and [11] in which loudspeaker compensation filters to simulate free-field listening are obtained using a two level spectrum with phase designed for minimum peak factor.

Beside arguing that ETFE with the choice of certain input signals is a viable alternative to MLS/Golay methods, one of the goals of this paper is to provide details of the experimental setup and procedure to aid researchers who want to establish their own HRTF measurement setup. Also, we want to alert the scientific community of the existence of our HRTF database (seven subjects to date). Although not as extensive as, e.g., the CIPIC database [3] which contains 45 subjects, our HRTFs have the advantage that they are measured at points spaced more uniformly around the subject. This is beneficial for fitting and interpolation based, for example, on spherical harmonics [12].

2. EXPERIMENTAL PROCEDURE

To measure the acoustical transfer functions or HRTFs we produce an input signal (sound broadcasted from a given space location) and measure the output of the system (conditioned microphone signals received at the entrance of the left and right ear canals). But HRTFs vary depending on the angle of incidence of the sound stimulus. Hence, it is necessary to present excitation signals in all directions where the HRTF ought to be measured. To this purpose, a set of loudspeakers (8 ohms Realistic 3" midrange tweeter, 700-20000 Hz) mounted on a semicircular, horizontally-rotating hoop, broadcast the acoustic signals used in the measurements (see Figure 1). To avoid excessive reflections only 6 loudspeakers are placed simultaneously on the hoop and recordings are taken in several sets to cover all desired azimuths. For each configuration of loudspeakers the hoop steps through all needed elevations automatically controlled by the computer.
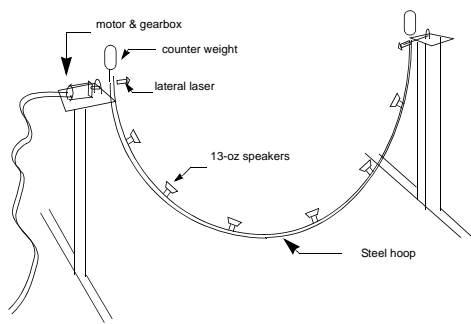

Figure 1: *Rotating semicircular hoop for HRTFs' measurement.*

HRTFs are measured at 1132 points around the head over a 5-degree double polar grid. The angular position of the loudspeakers from the center of the hoop defines the azimuth ϕ, while the angle of rotation of the hoop relative to the horizontal plane defines the elevation θ of the acoustic source. Position $(\phi, \theta) = (0, 0)$ deg corresponds to a source in front of the subject. The measurement grid is determined by the intersection of parallel vertical circles (described by the loudspeakers during the rotation of the hoop) with similar imaginary horizontal circles as shown in Figure 2.
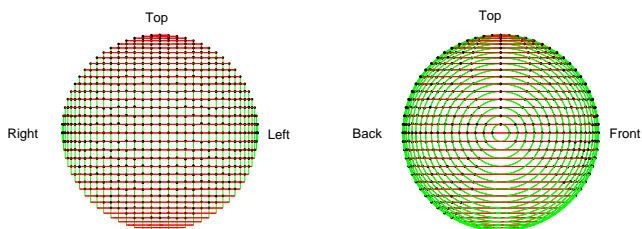

Figure 2: *The measurement grid is described by the intersection of horizontal and vertical circles.*

Signal production, motor control and signal acquisition are all integrated in a single MATLAB program running in a PC computer. This offers great benefit since all data are available on the spot, ready for post-processing, which is also done in MATLAB to take full advantage of its computational and graphical tools.

2.1. Apparatus

Data collection is performed in a sound attenuating room (single wall BioAcoustics 1.8 x 2.4 x 2.0 m) using a blocked ear canal technique. To dampen sound reflections the room walls are coated with dispersive foam (4.5 cm egg crate foam). The stainless steel hoop is rotated to specific positions by a motor (Animatics NEMA23) used in conjunction with a gearhead (100:1 NEMA 23) to satisfy the torque requirements. The motor control parameters (position, velocity, acceleration) are communicated to the motor via serial port. Adjustable plumbing clamps, attached to the back of each loudspeaker with screws, can be loosened or tightened to move the speakers along the hoop in steps of 5 deg.

2.2. Subject Preparation

Miniature microphones (Knowles FG3329) are inserted into the subject's ears flush with the entrance of the ear canal. The microphones are surrounded by a layer of swimming silicon plugs embedded into silicon moulding material which hardens within minutes conforming to the shape of the ear canal and sealing it. The subject's head is positioned at the center of the hoop, with the interaural axis coinciding with the axis of rotation of the hoop. Head position is monitored with a set of lateral lasers pointing to the entrance of the ear canal. Another laser, placed on top of the sujcet's head, is pointed forward onto a solar panel and is used to monitor head rotation throughout the experiment. If head misalignment is detected, the experiment stops automatically and a beep notifies the subject. After repositioning him/herself, the subject uses a reset switch to continue the experiment. Further details about the experimental setup and subject preparation can be found in [13].

2.3. Collecting the Acoustic Data

Figure 3 depicts the operation of the equipment used to broadcast acoustic stimuli and to record the conditioned microphone signals. The excitation signals are generated in MATLAB, converted to analog through a D/A port of a NI PCI-6071E card, power amplified and directed to one of the six loudspeakers through a demultiplexer (MAX388).

After collecting data from the subject, the input (reference) ꜱp at $\theta =$ cessary to ꜱ recorded
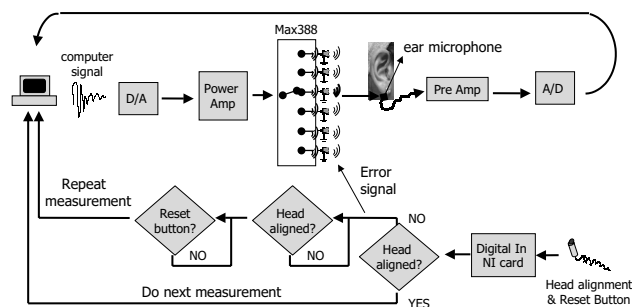

Figure 3: *Operational diagram of the HRTF setup.*

with the microphones are decoupled, amplified, and low-pass filtered with custom preamplifiers (4th order Bessel filter, cut off frequency of 18 kHz) [14] to avoid aliasing. Signals are sampled at a

rate of of 83.3 ksamp/sec and acquired into MATLAB through two A/D ports of the NI card, and they can be monitored on the computer screen. Although in theory the Nyquist sampling rate, i.e., twice the bandwidth of the system or 40 kHz, should be adequate, in practice, it is desirable to sample the data even at 6-10 times the system bandwidth [9].

To increase the signal to noise ratio (SNR), twenty-five pulses, originating from repetitive play of the same deterministic signal, are broadcasted and recorded for each instance. Pulses are 3.6 ms in duration with a repetition period of 43 ms to prevent overlap with echoes from the equipment and surrounding walls.

Figure 4 depicts three kinds of computer generated signals explored in obtaining the HRTFs: linear FM sweep (300 Hz-20 kHz), double-clock 7th order PRBS, and deterministic band noise (realization of white noise filtered with a 6th order Butterworth filter, 300 Hz-20 kHz). The actual signals delivered to the system are somewhat different and of longer duration due to loudspeakers distortion. Testing different signals allows us to examine the linearity of the HRTFs and to test what signals are more appropriate for the measurements. Signals like sweeps and PRBS which have a low crest factor, i.e., contain more signal energy for a given peak level, should be particularly favorable if the attainable SNR is limited by saturation and other large-signal nonlinearities of the system to be identified.
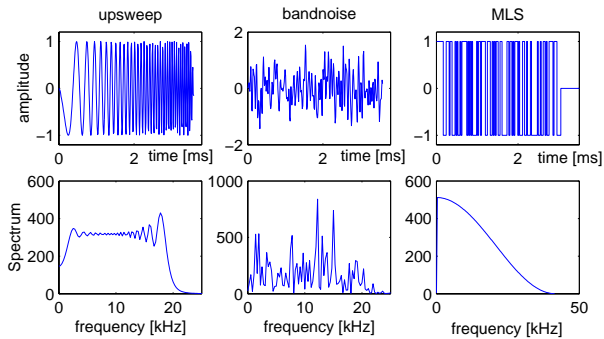

Figure 4: *Top: Computer generated excitation signals explored in the measurement of HRTFs. Bottom: their spectrum.*

3. EXTRACTION OF HRTFS FROM RAW DATA

The raw input and output data are preprocessed to keep only the sound recordings in a window where there is no echo, removing the (noisy) silent intervals between signals as well. This shortens processing time and makes the signal content less contaminated by noise. The window is 540 pts (6.5 msec) wide which limits the HRTF resolution to 154 Hz. The twenty-five pulses recorded for each instance differ only by random, zero-mean noise. Thus, averaging them coherently in time reduces the noise variance by a factor of 25 and improves SNR (see Figure 5). Good coherence among pulses is ensured since signal production (D/A) and sampling (A/D) operate with the same clock signal.

From the time-averaged pulses, the direction dependent transfer functions are computed by ETFE by taking the ratio of the Fourier transform of the acquired signal to the Fourier transform of the input signal. For each ear, these calculations result into a set of HRTFs, which depend on the position (ϕ, θ) of the acoustic
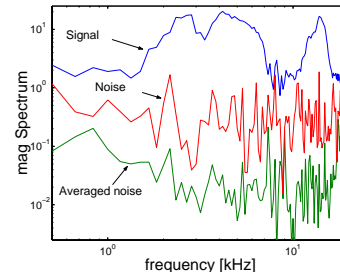

Figure 5: *Noise reduction achieved by time averaging of 25 pulses.*

source:

$$\mathrm{H}_L(f; \phi, \theta) = \frac{\mathcal{F}[\bar{y}_L(t; \phi, \theta)]}{\mathcal{F}[\bar{u}_L(t)]}; \quad \mathrm{H}_R(f; \phi, \theta) = \frac{\mathcal{F}[\bar{y}_R(t; \phi, \theta)]}{\mathcal{F}[\bar{u}_R(t)]}$$

where $\mathcal{F}[\cdot]$ is the discrete-time Fourier transform operator (here a zero-padded 1024 pt FFT), $\bar{y}_L = \sum_{i=1}^{25} y_{L,i}(t; \phi, \theta)$ is the time-averaged pulse for source located at (ϕ, θ), and the sub-indexes L and R refer to left and right, respectively. Likewise, the average reference pulse for the corresponding loudspeaker is $\bar{u}_L = \sum_{i=1}^{25} u_{L,i}(t)$.

The calculated HRTFs are only valid in the range of 700 Hz to 18 kHz. This limitation stems from the band pass characteristics of the loudspeaker (low cut-off frequencies of 700 Hz), and the microphone preamplifiers (cut-off frequency of 18 kHz). Therefore, to make the HRTFs suitable for sound rendering, they need to be extrapolated to reasonable values outside this range. The following approximations are applied: 1) The high frequency end is tapered to zero using a half Blackman window, starting at 18 kHz and reaching zero at 21 kHz. 2) At low frequencies the filter is approximated as a delay line. We do this by using linear interpolation between the computed and the ideal HRTF. The value of the delay for each HRTF is based on the measured time of arrival of the pulses obtained by threshold detection.

4. RESULTS AND DISCUSSION

In a preliminary experiment, run in a similar setup, the HRTFs of a mannequin head are obtained using the input signals discussed in Section 2.3. Figure 6 shows the right transfer function corresponding to (ϕ, θ)= (20,-5) deg. Except for occasional ouliers, the results are in very good agreement. Figure 7 shows a similar comparison of actual human HRTFs obtained in our current setup.

Our experience shows that sweeps tend to produce smoother HRTFs and at present we only employ sweeps in the measurements. One interesting point to notice is that the occasional spikes in the HRTFs extracted with band noise and PRBS appear randomly in different points and if different signals were used in the measurements, one could seek to reduce the variance and eliminate the occasional mismatch of the HRTFs with an algorithm that removes the outliers. A possible cause of spikes are the nonlinearities occurring in the microphones preamplifiers, e.g., slew-rate limits which could be excited by steep signal changes like the ones occurring in PRBS. A feasible remedy would be to smooth the PRBS signals with a low pass filter of 18 kHz such that steep changes are avoided. The variance of the ETFE is inversely proportional to $|\mathcal{F}[\bar{u}(t)]|^2$ [9]. Thus, low energy of the input signal
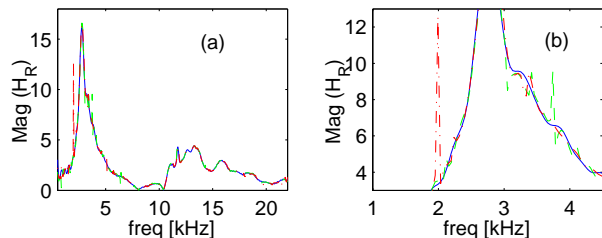
Figure 6: *Agreement of HRTF of a mannequin head obtained by three different excitation signals: (-) downsweep, (- -) PRBS, and (-.) band-noise. (a) right HRTF for (ϕ, θ)= (20,-5) deg. (b) zoomed in version of (a) to put in evidence the occasional spikes in the HRTFs obtained with PRBS and band noise input.*
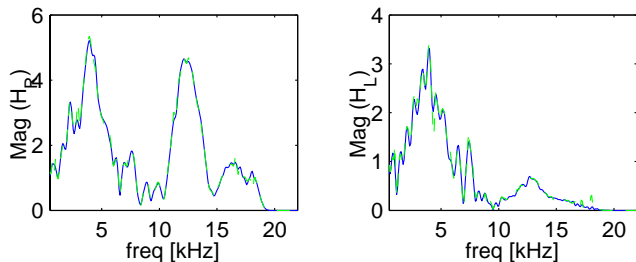

Figure 7: *Comparison of human blocked-canal HRTFs (ϕ, θ)= (20,-5) deg obtained by (-) upsweep and (- -) band noise excitation showing very good agreement.*

at some frequencies, increases the variance at that particular frequency which could also result in spikes.

It should be noticed that PRBS are not used here to facilitate the deconvolution of HRTFs, as it is customarily done with MLS and Golay methods. A reason for this is that loudspeakers tend to show nonlinear response when driven with this kinds of signals and to minimize nonlinearities requires to keep a low signal, thus compromising the achievable SNR.

As mentioned earlier, the computer-generated excitation signals shown in Figure 4 suffer some distortion and the spectrum of the actual acoustic signals differs from the ones shown in the figure. However, as long as the effective acoustic signals have a relatively smooth spectrum with energy well above signal at all frequencies of interest, signal distortion (linear or nonlinear) before reaching the ear should not be of concern in the ETFE.

5. CONCLUSIONS

In view of today's computational power and storage media, the advantages that MLS/Golay methods may have once enjoyed (speed of processing and low memory and circuitry requirements) have become obsolete. With ETFE post-processing of the row data is done off line and only takes about 12 min for the full data set, running in a PC Pentium 4, 1.8 GHz, with the much added flexibility of having the raw data available if one desires to explore different kinds of processing.

The ETFE avoids problems that may arise due to nonlinearities in the transformation path from computer signals to the actual acoustical signals. Linear transformation introduced by mi-

crophone preamps are cancelled out since they are also contained in the reference signal. However, ETFE is not immune to nonlinearities in the path after the signals are recorded by the microphones. Hence, it is important that the microphone preamplifiers have adequate slew rate and signal range.

The set of HRTFs (seven subjects to date) is available to the public and we hope that researchers who want to develop their own experimental setup will find the details presented here useful.

6. REFERENCES

[1] W. M. Hartmann, "How we localize sound," *Physics Today*, pp. 24-29, Nov. 1999.

[2] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman, "Localization using noindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 111–123, July 1993.

[3] V.R. Algazi, R.O. Duda, D.M. Thompson and C. Avendano, "The CIPIC HRTF Database," in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 2001.

[4] J.C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Amer.*, vol. 106, pp. 1480–1492, 1999.

[5] D. Zotkin, R. Duraiswami, and L. Davis, "Rendering localized spatial audio in a virtual auditory space," *to appear in IEEE Trans. on Multimedia*.

[6] D.D. Rife and J. Vanderkooy, "Transfer function measuremets with maximum-length sequence," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, June 1989.

[7] P. Zahorik, "Limitations in using Golay codes for head-related transfer function measurement," *J. Acoust. Soc. Amer.*, vol. 107, no. 3, pp. 1793–1796, March 2000.

[8] J. Vanderkooy. "Aspects of MLS Measuring Systems," *J. Audio Eng. Soc.*, vol. 42, no. 4, pp. 219–231, April 1994.

[9] L. Ljung *System Identification: Theory for the User*, second edition, Prentice Hall, New Jersey, 1999

[10] S. Mehrgard and V. Mellert, "Transformation characteristics of the external human ear," *J. Acoust. Soc. Amer.*, vol. 61, no. 6, pp. 1567–1576, June 1977.

[11] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Amer.*, vol. 85, no. 2, pp. 858–867, Feb. 1989.

[12] M.J. Evans, J.A.S. Agnus, and A.I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 104, no. 4, pp. 2400–2411, October 1998.

[13] J.K. Tulsi, *Using head related transfer functions for enhancing human elevation localization*, master thesis, Dept. of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA, December 2002.

[14] "Active filter design techniques," Chapter 16, excerpted from *Op Amps for Everyone*, literature no. sloa088, Texas Instrument, available at http://www-s.ti.com/sc/psheets/sloa088/sloa088.pdf