AES

Audio Engineering Society

Convention Paper

Presented at the 122nd Convention
2007 May 5–8 Vienna, Austria

Multiple Exponential Sweep Method for Fast Measurement of Head Related Transfer Functions

PIOTR MAJDAK[1], PETER BALAZS, and BERNHARD LABACK

[1] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
piotr@majdak.com
peter.balazs@oeaw.ac.at
bernhard.laback@oeaw.ac.at

ABSTRACT

Presenting sounds in virtual environments requires filtering of free field signals with head related transfer functions (HRTF). HRTFs describe the filtering effects of pinna, head, and torso measured in the ear canal of a subject. The measurement of HRTFs for many positions in space is a time-consuming procedure. To speed up the HRTF measurement the multiple exponential sweep method (MESM) was developed. MESM speeds up the measurement by interleaving and overlapping sweeps in an optimized way and retrieves the impulse responses of the measured systems. In this paper the MESM and its parameter optimization is described. As an example of an application of MESM, the measurement duration of a HRTF set with 1550 positions is compared to the unoptimized method. Using MESM, the measurement duration could be reduced by a factor of four without a reduction of the signal-to-noise ratio.

1. INTRODUCTION

1.1. Head Related Transfer Functions

A head related transfer function (HRTF) describes the sound transmission from the free field to a place in the ear canal [1] in terms of a linear time-invariant system. HRTFs depend on the angle of incidence and are highly subject dependent. A set of HRTFs, measured for different positions, can be used to create virtual free-field stimuli allowing access to virtual environment by presenting them via headphones [2].

The measurement of a HRTF can be performed by presenting a signal via a loudspeaker and recording the signal from a microphone, placed in or at the entrance of the ear canal. Then the recorded data is processed to obtain the HRTF. The measurement procedure of a single HRTF can be described as a process of system identification and must be separately performed for all positions of interest to get a complete set of HRTFs.

Generally, the HRTFs are independent of distance for sources beyond 1m [3]. This allows a measurement setup with two variables only: azimuth and elevation angles (Fig. 1). Arranging the loudspeaker along a given azimuth and elevation can be achieved in several ways. One method uses one speaker, which is moved on a fixed arc to the measured position keeping the subject in the center of the arc [4]. Another approach requires as many speakers as elevations measured, which allows to keep them at a constant position and turn the subject to set up required azimuthal position [1]. The mechanical requirements on the measuring system are much lower for the second method but it requires an implementation of a multichannel audio system. Fortunately, this became affordable nowadays.

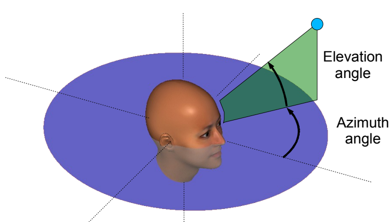

Figure 1. *Spatial setup of a HRTF measurement. The circle shows a sound source at a position specified by elevation and azimuth angles.*

The required spatial resolution of a HRTF set depends on the application field and is limited by the spatial localization ability of humans [5, 6]. Generally, it is necessary to cover at least the upper hemisphere where the resolution should be smaller than five degrees in the horizontal plane and at least ten degrees in elevation [7, 8]. Following this rule the amount of HRTFs in a set exceeds 1000 positions for the upper hemisphere. This may lead to a very time-consuming procedure because the total measurement duration strongly depends on the system identification method applied for the measurement of a single HRTF. Thus, it should be carefully chosen, taking different aspects into account.

1.2. Choice of the System Identification Method

Several issues should be considered choosing a system identification method for acoustical systems like HRTFs. The measurements are performed in noisy rooms, where a background noise reduces the signal-to-noise ratio (SNR) of the measurement. Also, the equipment, especially the power amplifier, adds noise to the excitation signal. These problems can be tackled increasing the energy in the signal, which can be achieved in two ways. One obvious possibility is to raise the amplitude of the signal. Unfortunately, the maximum amplitude is limited by the comfortable level for the subjects and the range of linear characteristics of the equipment. The other way is to extend the signal in time, which can be achieved using another, longer signal, or by repetition of the measurement. Anyhow, taking into account that the subject must keep still during the total measurement, the measurement duration should be as short as possible.

In case of unavailability of an anechoic chamber, measurements are performed in a reverberant sound booth. Long reverberation time requires long measurement duration to avoid artifacts like time aliasing or truncation of the impulse response of the measured system.

In addition to the SNR and measurement duration two further aspects should be considered: immunity to nonlinear distortions and time variability. Presenting signals via a loudspeaker produces nonlinear distortions in the signal, mostly due to the saturation effects of the loudspeaker membrane and the nonlinearity of the gain characteristics of the power amplifier. Furthermore, due to these distortions the measurement chain amplifier-speaker-room-subject-microphone-amplifier has to be described as a weakly nonlinear system. This can be modeled as a Volterra series with a small distortion factor [9]. The nonlinear artifacts are caused by the equipment and must be separated from the measurement results to obtain the HRTF. Due to the fact that the subject's head position may vary during the procedure, the measurement of HRTF must be considered as an identification of a weakly time-variant system. Thus, the method of choice should not produce any additional artifacts besides the changes in the HRTF due to head movements, which are unavoidable, even keeping the subject under surveillance with a head-tracking system.

Several system identification methods have been taken into consideration. The first one, periodic impulse excitation (PIE), uses a pulse train as excitation signal and is very easy to implement [10]. The signal has very low energy content with respect to the peak amplitude, which results in a high crest factor for long impulse responses. Thus, many repetitions have to be done to obtain a given SNR. Another procedure, dual-channel-FFT [11], uses white noise as excitation signal. It has similar disadvantages as the PIE. The family of binary pseudo random sequences help to increase the SNR, as they have a crest factor of 1, which is the lowest possible. Members of this family are the maximum length sequences (MLSs) which have been successfully applied for system identification [12]. Measurements with MLS are sensitive to nonlinear distortions. This problem can be reduced by lowering the amplitude of the signal [13], but, unfortunately, this results in lowering the SNR. The length of a MLS is 2^n-1 $(n\in\mathbb{N})$ samples, thus, the processing of longer sequences can be very time-consuming since the well known fast Fourier transformation (FFT) based on radix-2 algorithm can not be applied for the calculations. Golay-codes [14], another member of the family of binary pseudo random sequences, have a length of 2^n $(n\in\mathbb{N})$ samples, which allows usage of the FFT and thus faster processing. As a disadvantage, it has been shown, that measurements of HRTFs using Golay-codes can lead to non-negligible artifacts due to head movements [15]. The general problem of high sensitivity to nonlinear distortions remains unsolved using binary pseudo random sequences.

There is another signal family, which can be applied for system identification: frequency sweeps. Besides the measurement methods with linear sweeps such as time delay spectrometry [16] the system identification with exponential sweeps (ES) [17, 18] promises some interesting properties: separation of linear and nonlinear parts of weakly nonlinear systems, a low crest factor of $\sqrt{2}$ resulting in high SNR, fast processing using the FFT, deterministic signal generation, definition of measured frequency range, and low sensitivity to transient noise.

Measurements were performed using MLS and ES to obtain an overview of the system identification in our sound chamber (5.5 m x 5.5 m x 3 m semi-anechoic chamber, reverberation time: 80 ms, background noise level: 18 dB SPL) using our equipment (24-channel loudspeaker and microphone system, 48 kHz, 24 bit). The results showed that excitation signals (MLS or ES) with a length of about 1.5 s are required to achieve an SNR of at least 70 dB. Keeping in mind that a measurement of a HRTF set consists of over 1000 positions, this would lead to a total measurement duration of over half an hour. It is very hard for a subject to keep stock-still for such a long time. Thus, a way to speed up the measurement procedure is essential to perform HRTF measurements in our case.

One possibility to speed up the procedure is to measure several positions simultaneously respectively overlapping the measurements in time. Given that a multichannel equipment to drive different elevations simultaneously is available, all elevations at one azimuthal position, can be measured at once – provided an adequate system identification method can be applied. Under the prevailing circumstances the system identification with exponential sweeps was chosen to optimize the duration of HRTF measurement.

2. EXPONENTIAL SWEEP METHOD

System identification using exponential sweeps was described in detail in [18]. The exponential sweep (ES) is a sweep in frequency, see Figure 2, from ω_1 to ω_2 with the length of *T*:

$$x(t)=\sin\left[\omega_1\cdot\frac{T}{c}\cdot\left(e^{t\cdot\frac{c}{T}}-1\right)\right],\ \ t\in[0,T] \quad (1)$$

where c/T is the slew rate with $c=\ln(\omega_2/\omega_1)$.

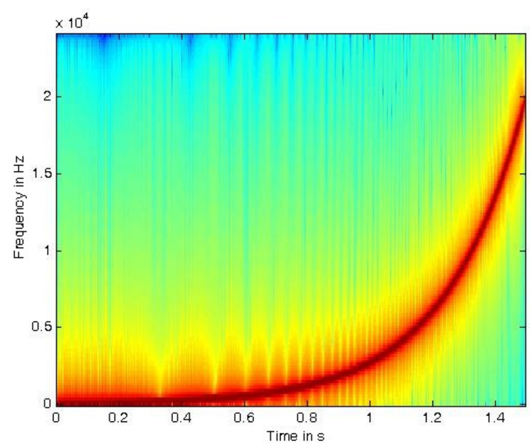

Figure 2. *Spectrogram of an exponential sweep from 50 Hz to 20 kHz with a duration of 1.5 seconds*

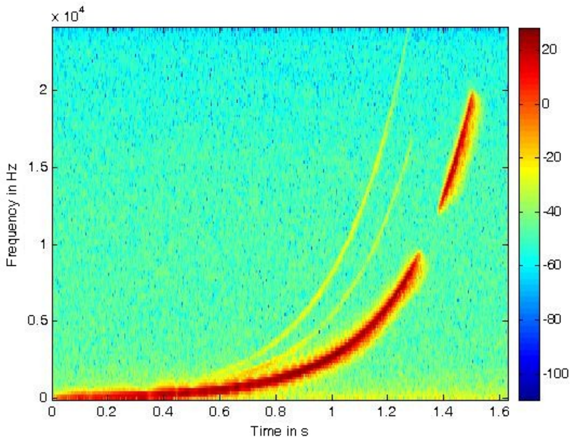

Figure 3. *Spectrogram of the output signal of a weakly nonlinear system, excited with the ES from Fig. 2. The linear part of the system was designed as a notch filter at 12 kHz to point out the changes in the system response.*

Applying the ES to a weakly nonlinear system produces an output signal $y(t)$ with higher order harmonics in addition to the linear part, see Figure 3 as an example.

The output signal is deconvolved using an inverse[1] sweep signal $\tilde{x}(t)$ to obtain a series of impulse responses (IR, see Figure 4):

$$s(t)=y(t)*\tilde{x}(t) \tag{2}$$

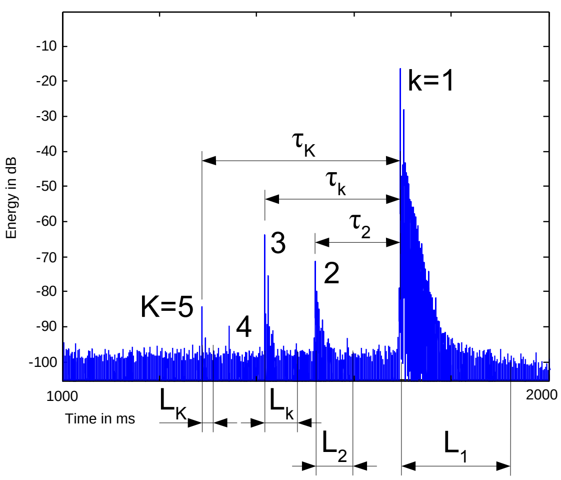

Figure 4. *Energy time curve of the deconvolved output signal $s(t)$. See text below for further details.*

The rightmost part of $s(t)$ (tagged with *k=1* in Fig. 4) represents the linear impulse response (LIR) with length L_1 of the linear part of the measured system. The remaining parts, left to the LIR, represent harmonic impulse responses (HIRs) of the nonlinear parts of the system. The HIRs are arranged in ascending order k from right to left (k from 2 to 5 in Fig. 4) where the exact distance τ_k (relative to the begin of the LIR) can be found in [17]:

$$\tau_k=\frac{T}{c}\cdot\ln k \tag{3}$$

For most weakly nonlinear audio systems, as in our case, the energy of a HIR decreases with increasing order and only a certain number K of HIRs can be identified in $s(t)$ ($K=5$ in Fig. 4). The parameter L_k is the lengths of the k-th HIR and depends on the reverberation of each part of the measured system. It is evident that the LIR and HIRs can be easily separated by windowing the interesting part of $s(t)$ if they are sufficient separated[2]. Then, the HRTF can be derived by the Fourier transformation of the LIR.

3. MULTIPLE EXPONENTIAL SWEEP METHOD (MESM)

The measurement of HRTFs at many elevations using N loudspeakers can be performed presenting the excitation signals almost simultaneously. The timing of the excitation of individual systems is essential to prevent superposition of the LIRs and HIRs, which would destroy the interesting parts of information. It can be derived using two mechanisms: interleaving and overlapping.

3.1. Interleaving of Sweeps

The aim is to interleave the LIRs and all HIRs of two systems in time. This is achieved by delaying the excitation of the second system in such a way that its LIR is placed between the linear and the second harmonic IR of the first system. This process can be extended to interleaving of as many systems as necessary, assuming a sufficient separation of their LIRs defined by:

$$\tau_2-L_2\geq(\eta-1)\cdot L_1\,. \tag{4}$$

Here η is the number of interleaved sweeps, L_2 is the length of the second order HIR of the first system, and L_1 is the maximum length of all LIRs of the other systems. These parameters are derived from the system identification of the individual systems (see Fig. 4). The separation given in Eq. 4 can be ensured by stretching the sweep signals to:

$$T'=\left[(\eta-1)L_1+L_2\right]\cdot\frac{c}{\ln 2} \tag{5}$$

[1] Inverse with respect to the convolution: $x(t)*\tilde{x}(t)=\delta(t)$

[2] The separation criterion depends on the requirements of the application.

An example of interleaving of four sweeps is shown in Figure 5. The measurement duration for a group of η interleaved systems is given by:

$$T_{grp}=T'+\eta\cdot L_1 \quad (6)$$

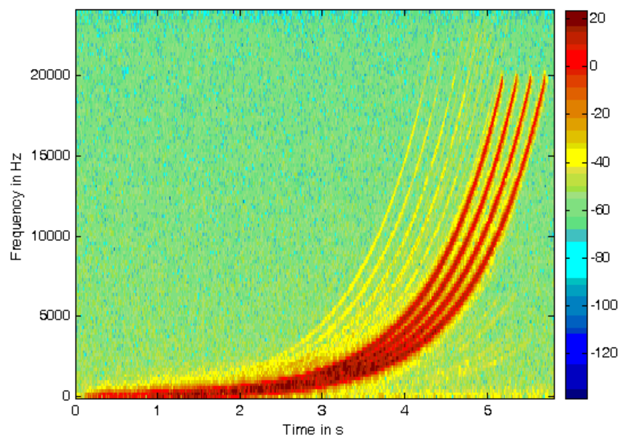

Figure 5. *Spectrogram of a recorded output signal as an example of interleaving of four sweeps.*

3.2. Overlapping of Sweeps

The overlap results from the fact that, given a small number of harmonics *K,* it is not necessary to wait to the end of the previous sweep when beginning the sweep. *K* can be derived analyzing the signal $s(t)$ as a result of the identification of an individual system (see Figure 4). The delay between two sweeps of τ_K+L_1 is sufficient enough to prevent superposing information of the LIRs of both systems, as shown in Figure 6.

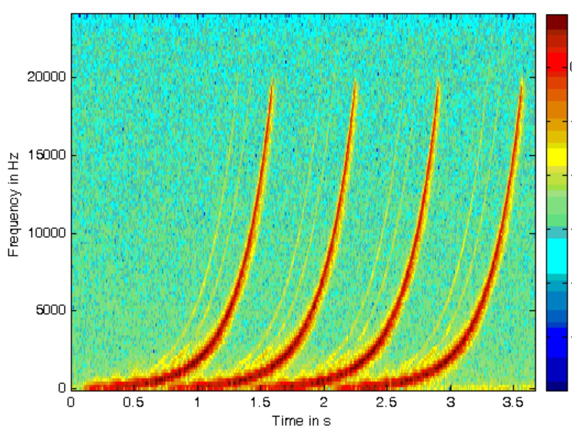

Figure 6. *Spectrogram of a recorded output signal as example of overlapping of four sweeps.*

3.3. Combining Interleaving and Overlapping

Both mechanisms, interleaving and overlapping are combined to form the multiple exponential sweep method (MESM). The overlapping mechanism is applied to N/η groups of interleaved sweeps with an intergroup delay of $\tau_K+\eta\cdot L_1$. Thus, the measurement duration of N systems is given by:

$$T_{tot}=T_{grp}+(N/\eta-1)\cdot(\tau_K+\eta\cdot L_1) \quad (7)$$

Once the number of interleaved sweeps is chosen, the excitation delay of the *i*-th sweep can be calculated by:

$$t_i=L_1\cdot(i-1)+\left\lfloor\frac{i-1}{\eta}\right\rfloor\cdot\tau_K \quad (8)$$

where $0\leq i\leq N$ and $\lfloor x\rfloor$ denotes the next lower integer of x. During the excitation via loudspeakers the microphone signal $y(t)$ is recorded, which is the sum of responses of the excited systems. Example of such an output signal is shown in Fig. 7.

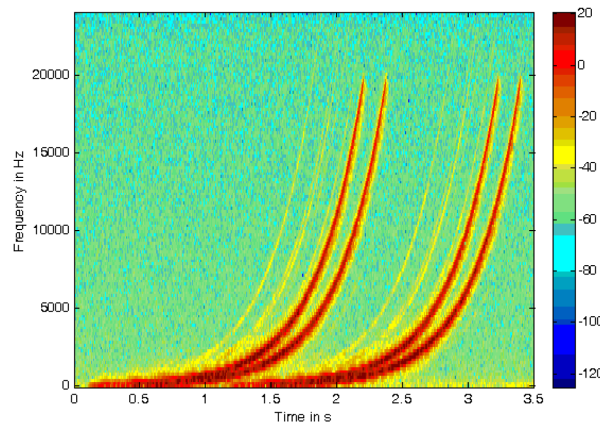

Figure 7. *Spectrogram of the recorded signal as an example of system identification with MESM (two groups of two interleaved sweeps) .*

Similar to the post processing of the exponential sweep, $y(t)$ is deconvolved to $s(t)$, which results in a series of LIRs and HIRs of all measured systems, as shown in Figure 8.

The separation of a LIR for a specific elevation is achieved by windowing $s(t)$. The shift of the window corresponds to the measured position and can be easily derived from t_i. Provided correct timing parameters, the LIRs can be separated without any artifacts. Notice that using MESM the HIRs partially overlap each other destroying the information about the nonlinear parts of the systems, which is not of interest in our case.

3.4. Comparison of Measurement Duration

The measurement durations for our measurement system with 22 loudspeakers were compared with the unoptimized ES method for different numbers of interleaved sweeps η. The length of each sweep was specified to be at least 1.5 s to fulfill the SNR requirements.

The frequency range was 50 Hz to 20 kHz. The measured reverberations were L_1=200 ms and L_2=20 ms, K was set to 13. These parameters were chosen after preliminary measurements, such that the described system represents an approximation of our HRTF measurement system.

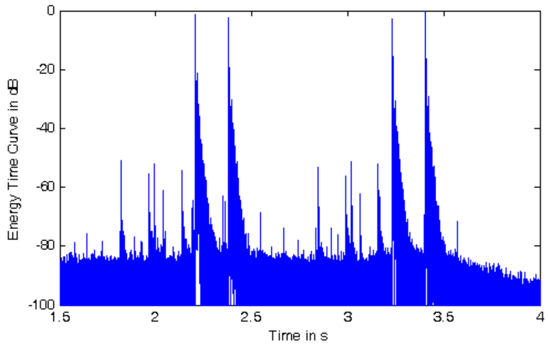

Figure 8. *Series of LIRs (higher peaks) und HIRs (lower peaks) of four weakly nonlinear systems. This figure corresponds to the output signal shown in Fig. 7. Notice that although the LIRs can be separated by windowing, the HIRs partially overlap each other.*

Using the ES method, the measurement duration for the given setup is 35.2 seconds. Applying MESM, i.e. overlapping seven groups of three interleaved sweeps leads to a duration of 7.103 seconds, which is 20% of the duration in the unoptimized ES case. Table 1 shows a comparison of measurement durations for different numbers of interleaved sweeps.

η	T' in s	T_{tot} in s	SNR gain [dB]
not optimized	1.5	35.2	-
1	< 1.5 [a]	12.162	0
2	< 1.5 [a]	7.729	0
3	1.815	**7.103**	0.8
4	2.68	8.119	2.52
6	4.408	9.766	4.68
12	9.595	13.94	8.06
22	18.239	20.44	**10.85**

Table 1. *Comparison of the measurement duration T_{tot} and SNR gain using MESM with different numbers of interleaved sweeps η. T' shows the sweep duration necessary for interleaving of η sweeps. The shortest measurement duration and the highest SNR are shown in bold.*

[a] According to Eq. 4, T' could be lower in this case, but it was kept to 1.5 s to fulfill the SNR requirements.

Instead of shortening the measurement duration, the MESM can be utilized to increase the SNR of the system identification. Interleaving of sweeps increases their lengths and hence their excitation energies, leading to a higher SNR of the measurement. Thus, a comparison of the SNR gains for different number of interleaved sweeps was included in Table 1. In this example, a SNR improvement of 10.85 dB could be achieved without any prolongation of the measurement.

4. SUMMARY

A new method for system identification based on the method of exponential sweeps [18] was introduced: the multiple exponential sweep method (MESM). This method is applicable for fast identification of multiple systems like HRTFs. MESM is adequate for measurements in noisy and reverberant rooms using weakly nonlinear systems and therefore, it offers advantages for measuring HRTFs in a simple sound chamber using ordinary equipment.

The MESM was examined considering a 22- loudspeakers system. It could be shown that, in this example, the measurement duration for 22 HRTFs at different elevations was reduced from 35.2 seconds to 7.103 seconds, which is about 20% of the duration in the unoptimized case. Applying MESM to our measurement setup and performing a measurement of a HRTF set consisting of 1550 positions, the total measurement duration could be reduced from 41 to about 10 minutes. This corresponds to an improvement by factor of four.

Besides the optimization of the measurement duration, MESM can be applied to improve the SNR of a measurement by keeping the total measurement duration constant and increasing the duration of individual sweeps. As an example, an SNR improvement of about 10 dB was shown.

5. ACKNOWLEDGMENTS

The authors would like to thank Robert Höldrich for his suggestions on this work and Michael Mihocic for the technical support.

6. REFERENCES

[1] H. Møller, M. F. Sørensen, D. Hammershøi and C. B. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, **43**(5): 300-321, May 1995.

[2] M. R. Schroeder, "Integrated-Impulse Method Measuring Sound Decay without Using Impulses," *J. Acoust. Soc. Am.*, **66**(2): 497-500, Aug. 1979.

[3] D. S. Brungart, W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J Acoust Soc Am.*, **106**(3 Pt 1): 1465-79, Sep. 1999.

[4] E. H. A. Langendijk and A. W. Bronkhorst, "Fidelity of three-dimensional-sound reproduction using virtual auditory display," *J. Acoust. Soc. Am.* **107**(1): 528-537, Jan. 2000.

[5] M. Morimoto and H. Aokata, "Localization cues in the upper hemisphere". *J. Acoust. Soc. Jpn.* (E) 5, 165-173, 1984.

[6] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners", *J. Acoust. Soc. Am.* **87**(5): 2188-2000, May 1990

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

[8] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.* 97:(6), pp. 3907-3908, June 1995

[9] M. Schetzen, "The Volterra and Wiener Theories of Nonlinear Systems," Robert Krieger Publishing, 1980

[10] J. M. Berman and L. R. Fincham, "The Application of Digital Techniques to the Measurement of Loudspeakers," *J. Audio Eng. Soc.*, vol. 25, pp. 370-384, June 1977.

[11] S. Müller and P. Massarani, "Transfer-Function Measurement with Sweeps," *J. Audio Eng. Soc.*, **49**(6): 443-471, June 2001

[12] D. D. Rife and J. Vanderkooy, "Transfer-Function Measurement with Maximum-Length Sequences," *J. Audio Eng. Soc.*, **37**(6): 419-444, June 1989

[13] C. Dunn and M. O. Hawksford, "Distortion Immunity of MLS-Derived Impulse Response Measurements," *J. Audio Eng. Soc.*, Vol. **41**(5): 314-335, May 1993

[14] B. Zhou, D. M. Green, and J. C. Middlebrooks, ''Characterization of external ear impulse responses using Golay codes,'' *J. Acoust. Soc. Am.* **92**(2): 1169–117, Aug. 1992

[15] P. Zahorik, "Limitations in using Golay codes for head-related transfer function measurement," *J. Acoust. Soc. Am.* **107**(3): 1793-1796, March 2000

[16] J. Vanderkooy, "Another Approach to Time-Delay-Spectrometry," *J. Audio Eng. Soc.*, 34(7/8): 523-538, July/Aug. 1986

[17] A. Farina, "Simultanous Measurement of Impulse Response and Distortion with a Swept-Sine Technique", presented at the *108th Convention of the Audio Engineering Society*, Paris, France, Feb. 2000

[18] Müller S. and Massarani P., "Transfer-Function Measurement with Sweeps", *J. Audio Eng. Soc.*, **49**(6): 443-471, June 2001