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

PIOTR MAJDAK, *AES Member,* **PETER BALAZS,** *AES Member,* **AND BERNHARD LABACK**
(piotr@majdak.com) (peter.balazs@oeaw.ac.at) (bernhard.laback@oeaw.ac.at)

Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

Presenting sounds in virtual environments requires filtering 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 overlapping sweeps in an optimized way and retrieves the impulse responses of the measured systems. MESM and its parameter optimization are described. As an example of an application of MESM, the measurement duration of an 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.

0 INTRODUCTION

A head-related transfer function (HRTF) describes the sound transmission from the free field to a place in the ear canal in terms of a linear time-invariant system [1]–[4]. HRTFs contain spectral and temporal cues, which vary according to the sound direction. A set of HRTFs measured for different positions can be used to create virtual free-field stimuli [5]–[7]. It has been shown that the directional-specific features differ among listeners [8], [9]. Subjects listening to HRTFs of other subjects show lower localization ability and report a less realistic environment [6], [7], [10], [11]. Thus measuring individual HRTF sets for each subject is necessary for most studies on localization in virtual environments.

The required spatial resolution of an HRTF set depends on the application field and is limited by the spatial localization accuracy of humans. Results of several studies (such as [12], [13]) imply that the spatial resolution of an HRTF set should be smaller than 5° in the horizontal plane and smaller than 10° in the vertical plane. Following this rule the number of HRTFs in a set exceeds 1000 positions for the upper hemisphere. For example, Gardner and Martin measured at 710 positions [14], and Algazi et al. measured at 1250 positions [9]. Using interpolation methods reduces the required number of HRTFs. However, the required number of measurements still remains large [15], [16]. Minnaar et al. [17] showed that the number of HRTFs could be reduced from thousands to 1130 by using interpolated HRTFs in their study. Because of the large number of HRTFs, the measurement duration may take tens of minutes, depending on the facilities. During the total measurement process the subject must keep still to avoid artifacts caused by head movements. Thus it is difficult for subjects to keep still during long HRTF measurements. Reducing the measurement duration makes this task more comfortable for the subject and reduces the measurement artifacts, yielding more exact measurements.

Three methods have been applied successfully to position sound sources at different spatial positions during the HRTF measurement. The first method uses one loudspeaker for each position on the sphere. However, this method is impractical for measuring hundreds of positions. The second method uses only one loudspeaker, which is moved to the measurement position [11], [18]. When using this method, it is not necessary to move the subject. Obviously, for this method only one position can be measured at a time. Furthermore, the requirements on the facilities are rather high because of moving parts in the room. The third method uses as many loudspeakers as elevations measured. The loudspeakers are stationary, mounted on an arc surrounding the subject, therefore requiring tens of loudspeakers instead of hundreds. The subject sits on a turntable and is rotated to the required azimuthal position [2]. The fixed mounting of the loudspeakers allows lower requirements on the facilities than the moving-loudspeaker method. However, using this method, the measured data may be affected by reflexions from the adjacent loudspeakers, which should be consid-

*Manuscript received 2006 December 4; revised 2007 April 24.

ered in the choice of the equipment. If loudspeakers are driven by a multichannel audio system, multiple sounds can be played simultaneously. Using this method the measurement of HRTFs for all vertical positions could be performed at once, which would speed up the measurement. Following this idea this study presents a system identification method, which allows almost simultaneous excitation of multiple systems and provides the separation of particular HRTFs from the recorded signal.

The measurement of HRTFs can be described as a process of identification of an electroacoustic system. Fig. 1 shows the signal path for the computer-aided system identification of acoustic systems such as HRTFs. The signal is presented from a given position via a loudspeaker, which is driven by a digital-to-analog converter (DAC) and a power amplifier. In our setup one set of a DAC, amplifier, and loudspeaker is required for each vertical position. Acoustic waves propagate in the measurement room and are altered by the torso, head, and pinna of a subject. Microphones are placed in the ear canals of a subject and capture the arriving sound. Finally, the binaural signal is recorded via two pre amplifiers and two analog-to-digital converters (one for each ear). Depending on the system identification method the recorded data are processed in different ways to obtain the HRTF. Using a traditional system identification method, the measurement of HRTFs is performed successively for all positions of interest.

The goal of the present study was to adapt the system identification method of exponential sweeps [19] for the measurement of HRTFs with the aim of shortening the measurement duration. This new method is called "multiple exponential sweep method" (MESM). It can minimize the amount of head movements during the measurement and thus reduce the extent of measurement artifacts due to these movements. As an application of MESM, the optimization results for our HRTF measurement system are presented.

1 OVERVIEW OF COMMON SYSTEM IDENTIFICATION METHODS

Many issues have to be considered when choosing a system identification method for acoustic systems. When the measurements are performed in noisy rooms, the 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. This problem can be addressed by increasing the energy in the signal, which can be achieved in two ways. One possibility is to raise the amplitude of the signal. Unfortunately the maximum amplitude is limited by the linear range of the equipment. For measuring HRTFs, the subject's comfortable level is also a limiting factor. The other way to increase the SNR is to increase the duration of the signal. Assuming an uncorrelated noise, the SNR of the measurement depends on the duration of the excitation signal. Doubling the signal duration increases the SNR by 3 dB. Usually, for HRTFs, longer measurements are achieved by repetitions of short measurements.

Another reason for extending the excitation signal in time is the reverberation of the recording environment. This reverberation requires the measurement to be prolonged in order to avoid artifacts such as time aliasing or truncation of the impulse response (IR) of the measured system. Usually HRTF measurements are performed in an anechoic chamber, where the reverberation time is very short. If an anechoic chamber is unavailable, measurements are performed in a sound booth with nonnegligible reverberation. This requires excitation signals longer than the time period where the reverberation decreases to the noise floor. However, for measuring HRTFs, the subject must keep still during the total measurement. Thus in this case the duration of the excitation signal is a compromise between technical requirements and the subject's comfort.

In addition to SNR and measurement duration, two further issues should be considered—nonlinear distortions and time variability. Presenting signals via loudspeakers yields nonlinear distortions, mostly due to the saturation effects of the loudspeaker membrane and the nonlinearity of the gain characteristics of the power amplifier. Due to these distortions, the measurement chain amplifier–loudspeaker–room–ear–microphone–amplifier has to be described as a weakly nonlinear system. This can be modeled as a Volterra series with a small distortion factor [20]. The nonlinear artifacts are caused by the equipment and must be separated from the linear component to obtain the HRTF. Furthermore, due to the fact that the subject's head position may move during the identification process, the HRTF measurement must be considered as an identification of weakly time-variant systems. Small head movements are unavoidable, even if the subject is kept under surveillance with a head-tracking system. Thus the measurement method of choice should be robust against small changes of the measured system.

Several system identification methods were taken into consideration. The first one, the periodic impulse excitation (PIE), uses a pulse train as the excitation signal and is relatively easy to implement [21]. For long impulse responses the response signal has very low energy content with respect to the peak amplitude (such as a high crest factor), resulting in a low SNR. Thus many repetitions have to be done to obtain a sufficient SNR. Another procedure, the dual-channel FFT [19], uses Gaussian white noise as the excitation signal. Its disadvantages are a high

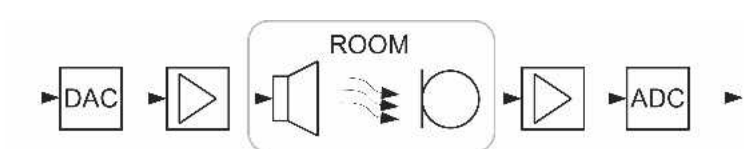

Fig. 1. Signal path for identification of systems such as HRTFs. Setup shown is for one channel only.

crest factor and its stochastic nature, which necessitates many repetitions of the measurement. The family of binary pseudorandom sequences [22] offers signals with a crest factor of 1. This allows measurements with the highest available SNR of signals with the same amplitude and duration. The most popular pseudorandom sequence is the maximum-length sequence (MLS). It has been applied successfully for system identification [23]. Unfortunately measurements with the MLS are sensitive to nonlinear distortions [24]–[27]. This problem can be reduced by lowering the amplitude of the signal [24], but this also results in lowering the SNR. The length of an MLS is always $2^n - 1 (n \in \mathbb{N})$ samples. Thus obtaining longer IRs from MLSs can be a very time-consuming process since the well-known fast Fourier transform (FFT) [28] based on the radix-2 algorithm cannot be applied for these calculations. Another type of binary pseudorandom sequences are Golay codes [29], which have a length of $2^n (n \in \mathbb{N})$ samples. They allow usage of the FFT and faster processing. As a disadvantage, it has been shown that measurements of HRTFs using Golay codes can lead to nonnegligible artifacts due to head movements [30]. The general problem of high sensitivity to nonlinear distortions remains unsolved using binary pseudorandom sequences [26], [27].

Another signal type, nonperiodic frequency modulated sweeps, can be applied for system identification [19], [31], [32]. In particular, the system identification with exponential sweeps (ES) shows some promising properties:

1) Separation of linear and nonlinear parts of weakly nonlinear systems
2) Low crest factor of $\sqrt{2}$ resulting in a high SNR
3) Fast processing using the FFT
4) Deterministic signal generation
5) Definition of the measured frequency range
6) Low sensitivity to transient noise, having an impact within a narrow frequency band only.

Thus the system identification with ES is a promising method for HRTF measurements. It was introduced by Farina [32] and was described in detail by Müller and Massarani [19] as a general system identification method for electroacoustic systems. MESM relies on the system identification with ES. Thus a short introduction to the ES method is given in the next section.

2 EXPONENTIAL SWEEP METHOD

The ES is a sweep in frequency from ω_1 to ω_2 with length T,

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

where c/T is the slew rate, with $c = \ln(\omega_2/\omega_1)$. An example of an ES is shown in Fig. 2.

A weakly nonlinear system excited by an ES produces a response signal $y(t)$ with higher order harmonics (see Fig. 3 as an example). To complete the system identification the output signal $y(t)$ is processed to $s(t)$,

$$s(t) = y(t) * x'(t) \tag{2}$$

where $x'(t)$ is the inverse[1] sweep. The inverse sweep compensates for the group delay and the magnitude spectrum of the excitation signal. It can be derived from the excitation signal by applying the inverse Fourier transform,

$$\boldsymbol{X}'(\omega) = \frac{\boldsymbol{X}(-\omega)}{|\boldsymbol{X}(\omega)^2|} \tag{3}$$

where $\boldsymbol{X}(\omega)$ is the complex spectrum of the excitation sweep $x(t)$ and $\boldsymbol{X}'(\omega)$ is the complex spectrum of the inverse sweep. Fig. 4 shows $s(t)$ as an example for the deconvolution of the response signal in Fig. 3.

[1] Inverse with respect to the convolution, $x(t) * x'(t) = \delta(t)$.

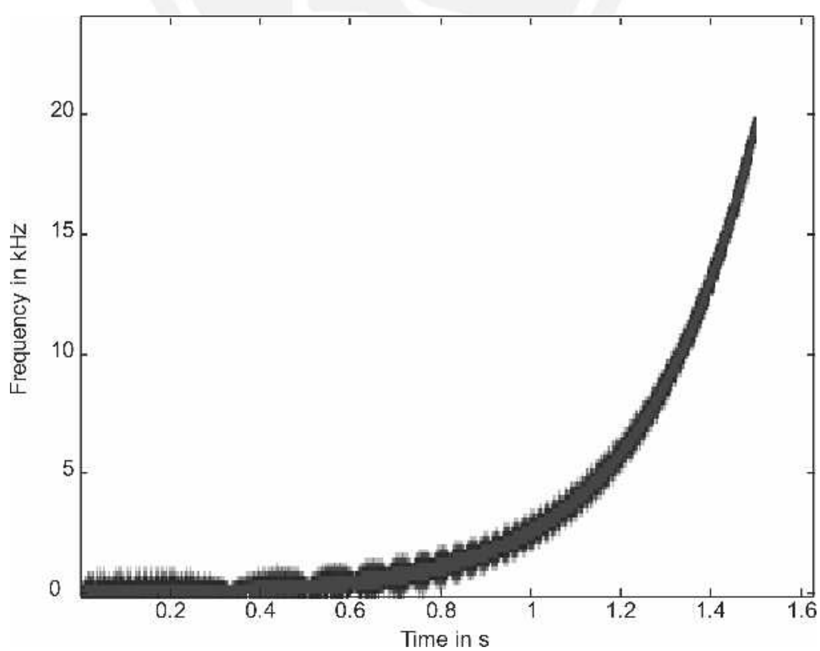

Fig. 2. ES spectrogram from 50 to 20 000 Hz, 1.5-s duration.

The result $s(t)$ is a series of harmonic impulse responses (HIRs). The HIRs are arranged in ascending order k from right to the left. The rightmost part of $s(t)$, beginning at $k = 1$, represents the first-order HIR, which is the IR of the linear part of the measured system. The length of the IR is given by L_1. The remaining parts, to the left of the IR, represent higher order HIRs of the measured system. The parameter L_k describes the duration of the kth HIR. The exact distance from the start of the kth HIR relative to the start of the IR is given by [32]

$$\tau_k = \frac{T}{c} \ln k. \tag{4}$$

For the most weakly nonlinear audio systems the energy of the kth HIR decreases with increasing order k, and only a certain number K of HIRs can be identified in $s(t)$. ($K = 5$ in the example from Fig. 4.) It is evident that if the IR does not overlap with other HIRs, it can easily be separated from the rest of $s(t)$ by windowing.

For a satisfactory system identification the sweep duration T must be well chosen. The choice must be made considering two aspects. First the necessary sweep duration depends on the SNR requirements and the noise conditions of the room. Assuming an uncorrelated noise, the SNR can be increased by 3 dB by doubling the sweep duration. Second the HIRs must not overlap each other. This can be achieved by choosing a sweep duration considerably higher than the reverberation time. In practice the sweep duration is chosen on the basis of trial and error. As a guideline, Müller and Massarani [19] describe some

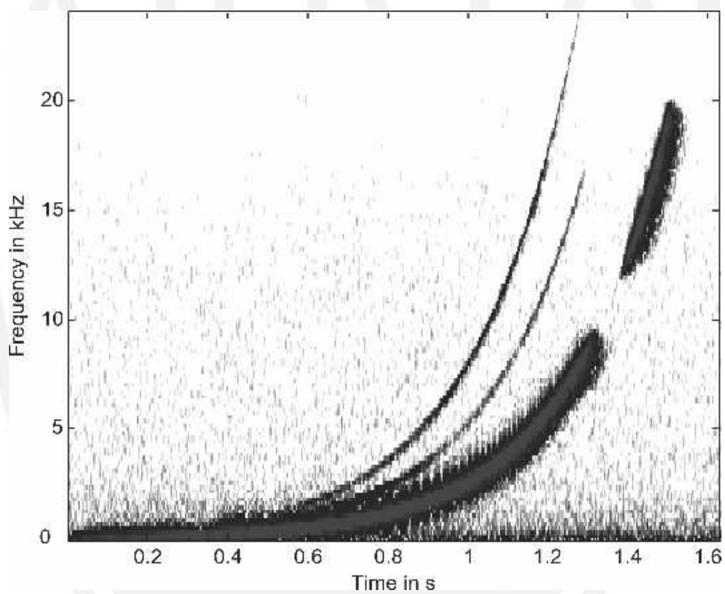

Fig. 3. Response-signal spectrogram of weakly nonlinear system excited with ES of Fig. 2. System is a notch filter with center frequency at 12 kHz and additional weak nonlinearity and noise.

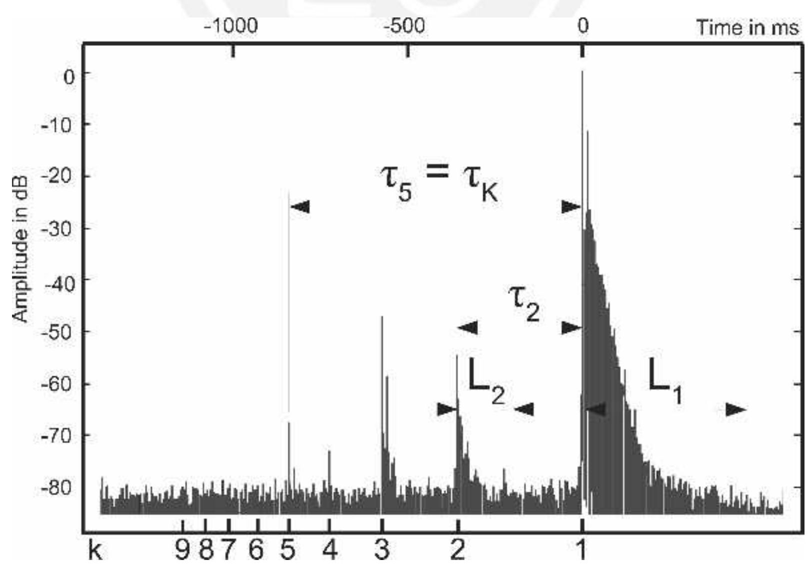

Fig. 4. Series of harmonic impulse responses (HIRs) of weakly nonlinear system as a result of deconvolution of response signal of Fig. 3. Data were normalized and are presented in dB.

effects of the sweep duration on the results in high-quality measurements of room IRs.

3 MULTIPLE EXPONENTIAL SWEEP METHOD (MESM)

The easiest, unoptimized method for the identification of multiple systems is to identify them system by system. This is achieved by playing the exponential sweep via a loudspeaker for the first system and recording the response signal. Then the procedure is repeated for each remaining system. Because of the reverberation additional length of L_1 must be considered in the recording duration for each system. For N systems this results in a measurement duration of

$$T_{ES} = (T + L_1)N. \tag{5}$$

In the MESM the sweeps overlap in time. This results in a shorter measurement duration of multiple systems. The timing of the excitation of the systems is essential to prevent a superposition of IRs with other HIRs. MESM provides the following:

1) Identification of the linear parts of weakly nonlinear systems
2) Correct timing of sweeps
3) Optimization of the measurement duration at a given SNR.

The derivation of the timing is based on the interaction of two mechanisms—interleaving and overlapping.

3.1 Interleaving

Consider the effect of applying a sweep to a weakly nonlinear system and, after a small delay, applying the same sweep to a second system. Recording the summed response signal and applying the deconvolution process will result in a signal where the HIRs of the two systems are interleaved in time. The interleaving mechanism results from delaying the excitation of the second system in such a way that its IR is placed between the IR and the second-order HIR of the first system. This process can be easily generalized to interleave HIRs of as many systems as necessary. Fig. 5 shows a response signal for the interleaving using four sweeps.[2] The result of the deconvolution of this signal is shown in Fig. 6.

The end of the second-order HIR and the beginning of the IR must be sufficiently separated to include all IRs. The separation is achieved by stretching the sweep to get $\tau_2 - L_2 > L_1(\eta - 1)$ (see Fig. 4), where η is the number of interleaved systems. Using Eq. (4) the sweeps must have a minimum duration of

$$T' = [(\eta - 1)L_1 + L_2]\frac{c}{\ln 2}. \tag{6}$$

In cases where T' could be smaller than the original duration T, the new duration T' must be set to T to fulfill the SNR requirements. With the new sweep duration T' the beginning of the kth HIR, relative to the beginning of the IR, is given by

$$\tau'_k = \frac{T'}{c}\ln k. \tag{7}$$

To apply the interleaving procedure in a measurement, the sweeps are played at the time $(i - 1)L_1$, where i is the

[2]Notice the thin curves in Fig. 5 to the right of the linear responses of the four systems. They represent the intermodulation components probably resulting from the nonlinearity of the microphone preamplifier. After the deconvolution the intermodulation components appear as peaks after the last IR (between 5.9 and 6.5 s in Fig. 6). Because of their low amplitude and good temporal separation from the IRs the intermodulation components do not affect the interesting parts of the data and can be ignored in this study.

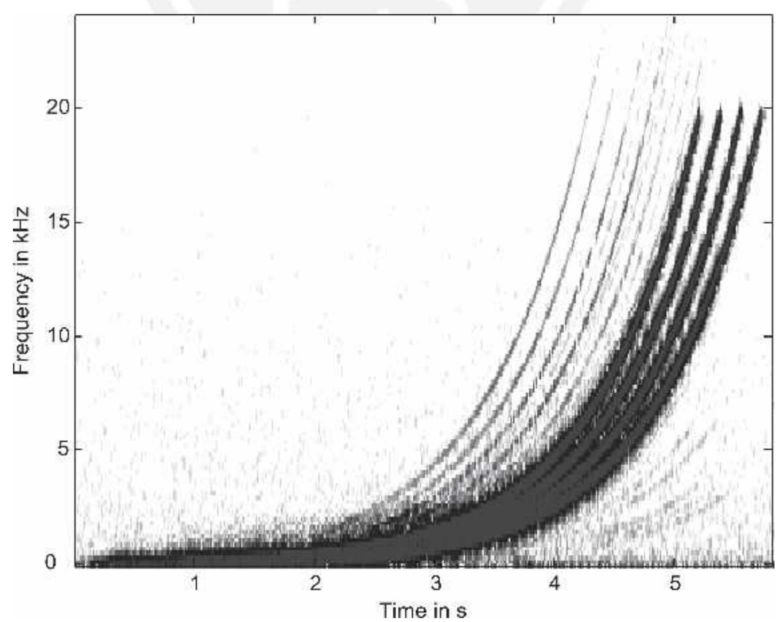

Fig. 5. Response-signal spectrogram as an example of interleaving using four sweeps.

index of the identified system, $0 < i \leq \eta$. The last sweep ends at $T' + (\eta - 1)L_1$. The recording of the response signal must be extended by L_1 at the end of the last sweep to capture the reverberation of the last system. Thus the measurement duration for a group of η interleaved systems is given by

$$T_{\mathrm{grp}} = T' + \eta L_1. \tag{8}$$

The measurement duration for N systems using the interleaving mechanism for η systems results in

$$T_{\mathrm{INT}} = \frac{N}{\eta} T_{\mathrm{grp}} = \frac{N}{\eta} T' + N L_1. \tag{9}$$

A comparison of Eqs. (9) and (5) shows that T' and η determine the change in the measurement duration. If their ratio is smaller than T then the interleaving mechanism results in an improvement in the measurement duration. An additional improvement can be achieved by introducing the overlapping mechanism.

3.2 Overlapping

In the most simple and straightforward method for the system identification of multiple systems, it is logical to play a single sweep, wait for its end, wait the length of the reverberation time, and then play the next sweep. However, in systems with a small number of harmonics, which is the case for weakly nonlinear systems, it is not necessary to wait for the end of the previous sweep. As long as the highest harmonic of the next sweep response does not interfere with the reverberation caused by the previous sweep the sweeps may overlap in time. An example of a response signal for four overlapping sweeps is shown in Fig. 7. In this example the second sweep begins already at 0.7 s while the first sweep is being played until 1.5 s. The deconvolution result of the response signal is shown in Fig. 8. It is evident that even though the sweeps overlap in time, the responses of the systems do not interfere with each other. Notice the difference in the interleaving mechanism: the HIRs of the particular systems are not interleaved in the overlapping mechanism.

If a sweep starts too early, the HIRs of the corresponding system overlap the IR of the previous system. In this case the information about the linear part of this system is destroyed. Thus a sufficient delay between two sweeps must be provided. The minimum delay preventing superimposing the information is given by

$$\Delta t_{\mathrm{OV}} = \tau_K + L_1 = \frac{T}{c} \ln K + L_1 \tag{10}$$

where K is the maximum number of harmonics found in $s(t)$. Overlapping N sweeps, the measurement duration is given by

$$T_{\mathrm{OV}} = T + (N-1)(\tau_K + L_1). \tag{11}$$

3.3 Combining Interleaving and Overlapping

Both mechanisms, interleaving and overlapping, are combined to form the MESM. First, interleaving for η systems is applied, resulting in N/η groups. Then these groups are overlapped with the intergroup delay of $\tau'_K + \eta L_1$. For a given η, the excitation time for the ith sweep is calculated by

$$t_i = L_1(i-1) + \left\lfloor \left(\frac{i-1}{\eta} \right) \right\rfloor \tau'_K \tag{12}$$

where $0 < i \leq N$ and $\lfloor x \rfloor$ denotes the next lower integer of x. During the excitation of the systems the response signal $y(t)$ is recorded. It represents the sum of the individual responses of the systems. Fig. 9 shows an example of such

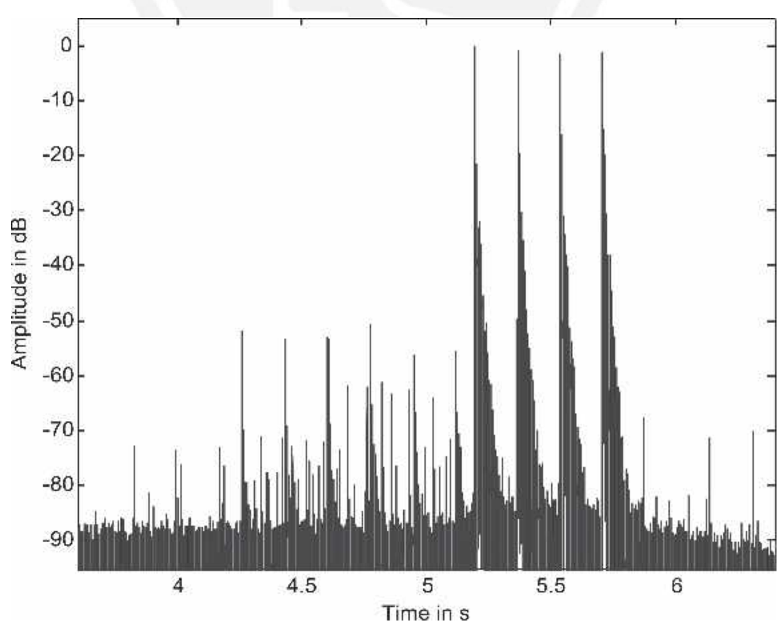

Fig. 6. Series of IRs (higher peaks) and higher order HIRs (lower peaks) of four weakly nonlinear systems. Data correspond to signal of Fig. 5. Series was normalized and is presented in dB. Time axis was adjusted to show important features of signal.

a response signal.[3] Applying deconvolution on this signal results in a series of HIRs, as shown in Fig. 10.

The extraction of a particular IR is done by windowing the signal $s(t)$. The shift of the window corresponds to the measured system and can easily be derived from t_i. Provided the timing parameters are correct, the IRs can be separated without any artifacts. The higher order HIRs partially overlap each other, destroying the information about the nonlinear parts of the systems. However, this is not relevant in cases where only the transfer functions are of interest.

The measurement duration is given by

$$T_{\mathrm{MESM}} = T' + \tau'_K\left(\frac{N}{\eta} - 1\right) + NL_1. \tag{13}$$

Comparing Eqs. (5) and (13) the term TN changed to $T' + [1 + (N - \eta)(\ln K/\eta c)]$. This shows that in MESM the measurement duration does not depend on the number of systems N only. It can be reduced for lower K values and/or for η values higher than 1. This means that according to Eqs. (13) and (5), the measurement duration finally depends on the parameters c, L_1, L_2, and K. The constant c is given by the start and end frequencies of the sweep.

[3]Notice the thin curves in Fig. 9 to the right of the linear responses of the two interleaved groups. They represent the intermodulation components described in footnote 2 and appear after the deconvolution as peaks (in Fig. 10 at 2.5 and 3.6 s). They interfere only with the higher order HIRs of the subsequent systems. Thus they have no relevance in this study.

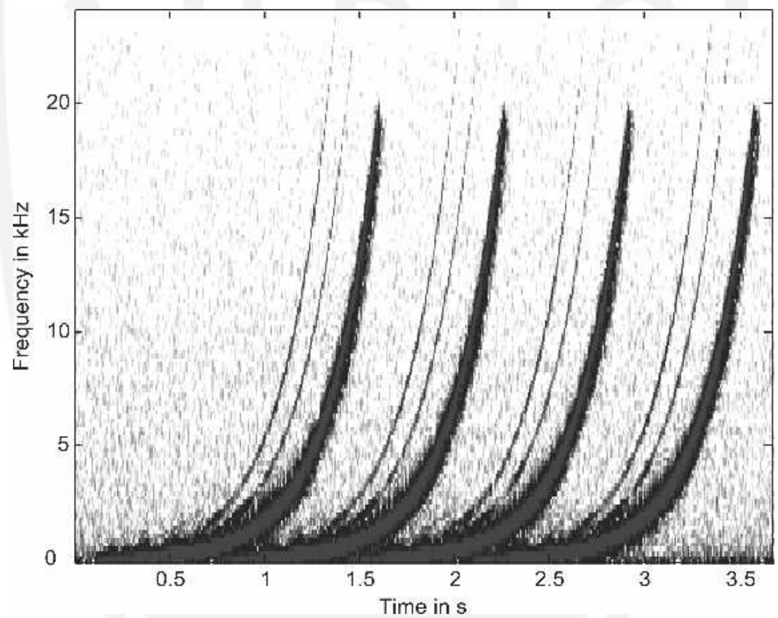

Fig. 7. Response-signal spectrogram as an example of four overlapped sweeps.

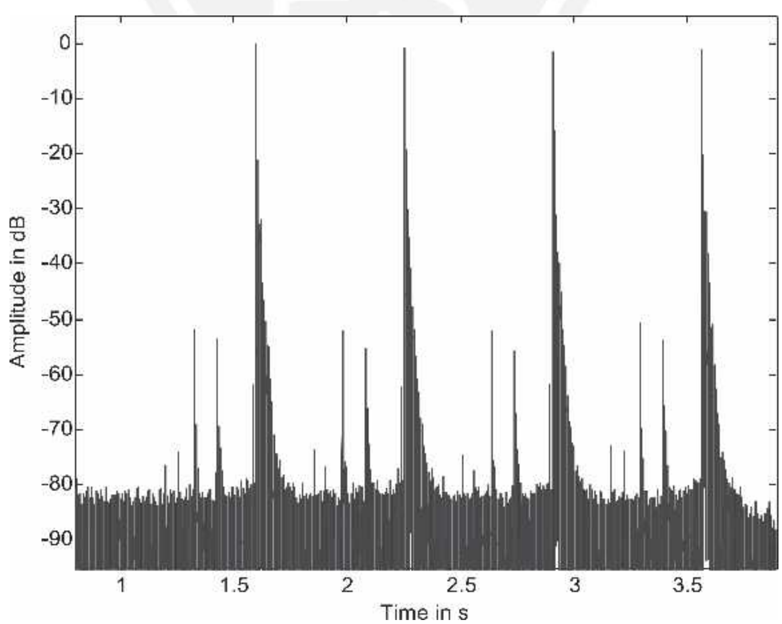

Fig. 8. Series of HIRs calculated from signal of Fig. 7. Data were normalized and are presented in dB.

The other parameters are derived from a preliminary measurement before using MESM, which is discussed in the next section.

3.4 Measurement Procedure

The values L_1, L_2, and K are required for the calculation of the MESM parameters—sweep duration T' and excitation time t_i. These parameters are obtained from the reference measurement performed prior to the first measurement with MESM. The reference measurement is done using an unoptimized method (such as the ES or MLS method) and should represent the baseline condition for further measurements with the same equipment. In the case of HRTF measurements, the reference measurement is performed in the absence of the subject, with microphones placed in the center of the loudspeaker arc. It is important that the retrieved parameters do not change noticeably when switching from the reference to the actual measurement of the systems. This can be assumed for the HRTF measurement, as outlined in the following. It is assumed that the HRTF measurement is performed in a semianechoic chamber. Placing microphones in the ear canals does not change the IR length noticeably because the duration of HRTF filters is short (several milliseconds) compared to the duration of the room IRs (up to several hundreds of milliseconds). Furthermore this does not add additional nonlinear components to the response signals because the HRTF is a linear filter. Thus the parameters

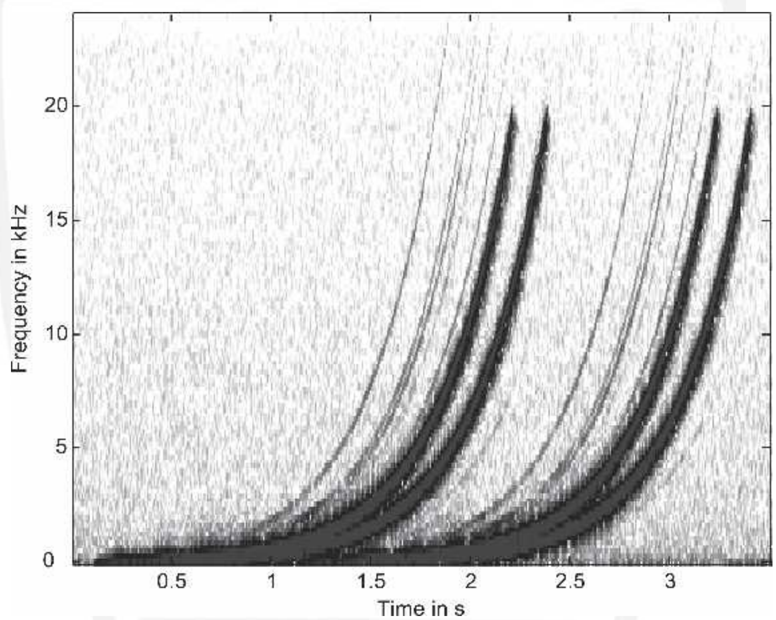

Fig. 9. Recorded-signal spectrogram as an example of a system identification using MESM (two groups of interleaving for two sweeps).

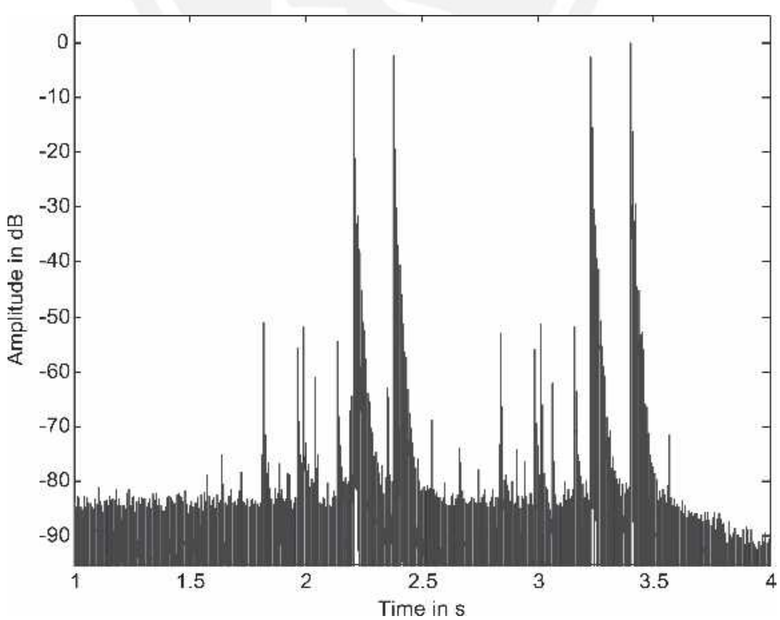

Fig. 10. Series of HIRs obtained from signal of Fig. 9. Data were normalized and are presented in dB. Notice that although IRs can be separated by windowing, HIRs partially overlap each other.

L_1, L_2, and K do not change noticeably from the reference measurement to the HRTF measurement.

The reference measurement results in HIRs for all audio channels, where the parameters, L_1, L_2, and K are derived for every individual channel. They may differ from channel to channel. Considering that the differences between channels correspond to different elevations, the following simplification allows to reduce the number of parameters. By changing the excitation position in the room, the occurrence of early reflexions in the room IR may vary, but the reverberation time does not change considerably.[4] This means that the parameters L_1 and L_2 are similar for all channels as they describe the reverberation time of the systems. Thus the maximum values of all channels for L_1 and L_2 can be used for MESM. The parameter K is the number of harmonics in the response signal and describes the extent of the nonlinearity of the equipment. Using the same type of amplifier and loudspeaker for all audio channels reduces the variability of K between the channels. In the case of a different K for every channel, the maximum value of K is used for optimization with MESM.

Once the values L_1, L_2, and K are known, the optimal number of interleaving and the timing of excitation are calculated. Then the measurements using MESM can be performed. Switching from the ES method to MESM one additional aspect must be considered—the amplitude of the excitation signals. Using MESM, multiple audio channels are used simultaneously. The sound waves created by different loudspeakers superimpose each other in the room. The microphones capture the summation result in the room (for the reference measurement) or in the ear canal (for the HRTF measurement). Depending on the number of overlapped sweeps, the amplitude of the summation signal is higher than the amplitude of the individual signals. This may result in clipping at a microphone preamplifier or at the analog-to-digital converter. Furthermore, for very high sound pressures, additional nonlinear components can be introduced by the microphone, causing changes in the parameter K. To avoid these problems sufficient headroom must be considered during the reference measurement. The following procedure is suitable to deal with the amplitude problem.

1) Perform the reference measurement using ES. Determine the sweep duration and excitation amplitude fulfilling the SNR requirements and allowing enough headroom for every audio channel.

2) Retrieve parameters L_1, L_2, and K from the HIRs. Use them to calculate the MESM parameters.

3) Perform the reference measurement using MESM. Adapt the amplitude to obtain enough headroom. The SNR requirement may not be met due to the lowering of the amplitude.

4) Repeat the reference measurement using ES. Check the SNR and extend the sweep duration if the SNR is too low. Do not change the amplitude.

5) Calculate the MESM parameters using L_1, L_2, and K from the first measurement and the new sweep duration from the latest measurement.

Assuming no changes in the equipment and room configuration this procedure is performed only once and results in the MESM parameters for all further measurements.

3.5 Optimization of Parameters

Assuming that the parameters L_1, L_2, K, and T are known, the MESM parameters are calculated, resulting in the measurement duration T_{MESM}. T_{MESM} can be optimized with regard to either of the two criteria—shortest measurement duration or highest SNR of the measurement.

In the first case the sweeps are overlapped and stretched to get the shortest possible measurement duration without lowering the measurement SNR. This can be achieved by finding the optimal number of interleaved systems $\eta_{\mathrm{opt},T}$ for which $T' \geq T$ and T_{MESM} is a minimum. To get $\eta_{\mathrm{opt},T}$, T_{MESM} is calculated by expansion of Eq. (13) with Eqs. (6) and (7),

$$T_{\mathrm{MESM}}(\eta) = \frac{1}{\ln 2}\left[\eta L_1(c - \ln K) + \left(\frac{L_2 - L_1}{c - \ln K} + NL_1 \ln 2K\right) + \frac{(L_2 + L_1)N \ln K}{\eta}\right]. \tag{14}$$

The solution of Eq. (14) with respect to η implies two cases,[5] $L_1 \geq L_2$ and $L_1 < L_2$, which must be handled separately. The first case is the usual case for weakly nonlinear systems where the reverberation of the linear part is longer than the reverberation of the second-order harmonic. In this case the solution for T_{MESM} has a negative pole at $\eta = 0$ and is a strictly monotonically increasing function. Thus for the shortest measurement duration $\eta_{\mathrm{opt},T}$ is the lowest η not breaking the requirement of $T' \geq T$. This is solved by using Eq. (6) to get

$$\eta_{\mathrm{opt},T} = \left\lceil \frac{T \ln 2}{cL_1} + \frac{L_1 + L_2}{L_1} \right\rceil_{L_1 \geq L_2} \tag{15}$$

where $\lceil x \rceil$ is a function that returns the next higher integer of x. The solution for the second case, $L_2 > L_1$, can be found by differentiation of Eq. (14). But this case is rather unusual for weakly nonlinear electroacoustic systems and is not handled explicitly here.

In the second optimization mode the SNR is maximized with the condition that the measurement duration does not

[4]The reverberation time of a room primarily depends on the sound absorption of the walls, floor, and ceiling. For a diffuse sound field, the position of the sound source has no effect on the sound absorption. In reverberant rooms the sound field becomes diffuse very fast because of a high density of reflections. Thus the reverberation time is assumed to be almost position independent.

[5]The solution of Eq. (14) requires $(c/\ln K) > 0$. This is the case if $K\omega_1 \leq \omega_2$. The opposite case, $K\omega_1 > \omega_2$, corresponds to cases where the frequency of a relevant harmonic in the response signal is higher than the end frequency of the sweep. Identification of nonlinear systems may lead to this. However, the convolution of such a response signal with the inverse sweep acts like a bandpass filter, cutting all frequencies below ω_1 and above ω_2. Thus in further analysis $K\omega_1 \leq \omega_2$ can always be assumed.

exceed the duration obtained with the unoptimized method. The SNR gain g represents the SNR relative to the SNR obtained with the unoptimized method and can be defined as

$$g = 10 \log_{10}\left(\frac{T'}{T}\right). \quad (16)$$

Following this definition, g increases monotonically with T'. Furthermore T' increases monotonically with η [see Eq. (6)], showing that g grows with η. Therefore the goal is to find the highest η fulfilling the condition $T_{\mathrm{MESM}} \leq T_{\mathrm{ES}}$. To achieve that, Eqs. (13) and (5) are combined to

$$T' + \tau'_K\left(\frac{N}{\eta} - 1\right) \leq NT. \quad (17)$$

After multiplication of Eq. (17) with η and substitution with Eqs. (6) and (7), the resulting quadratic inequality is solved, resulting in $\eta_{\mathrm{opt,SNR}}$, the optimal number of interleaved sweeps for highest g. The formula of the analytic solution exceeds the scope of this paper and can be obtained from the authors upon request. For a given $\eta_{\mathrm{opt,SNR}} < N$, T' results from Eq. (6). In cases where $\eta_{\mathrm{opt,SNR}} = N$, following Eq. (8), T' is given by

$$T' = T_{\mathrm{ES}} - NL_1. \quad (18)$$

With these optimization results the excitation time for each sweep is calculated using Eq. (12). Then the system identification can be performed following the procedure described in Section 3.4.

4 HRTF MEASUREMENT AS AN APPLICATION OF MESM

The MESM was used to speed up an HRTF measurement. The facilities included a semianechoic room with dimensions of 5.5 m × 5.5 m × 4 m and a reverberation time of 80 ms (at 150 Hz). A 24-channel digital audio interface (RME ADI-8) with a sampling rate of 48 kHz and a resolution of 24 bit was used. The A-weighted sound pressure level (SPL-A) of the background noise in that room was 18 dB (20 μPa) on a typical weekday. Twenty-two loudspeakers (custom-made boxes with VIFA 10 BGS as drivers; variations in the frequency response ±4 dB in the range from 200 to 16000 Hz) at fixed elevations from −30° to 80° were used. They were driven by amplifiers adapted from Edirol MA-5D active loudspeaker systems. The loudspeakers and the arc were covered with acoustic damping material to reduce the reflection from the adjacent parts. The total harmonic distortion of the loudspeaker–amplifier systems was on average 0.19% (at 63 dB SPL and 1 kHz). The in-ear microphones (Sennheiser KE-4-211-2) were connected via preamplifiers (RDL FP-MP1) to the digital audio interface. The microphones had low sensitivity to the acoustic signal due to their small diameters (4.75 mm). Thus a length of the ES (frequency range from $\omega_1 = 50$ Hz to $\omega_2 = 20$ kHz) of about 1.5 s was essential to achieve a SNR of at least 70 dB. For the reference measurement the left-ear microphone was placed in the center of the loudspeaker arc. The amplitude of the signals was adjusted to an SPL-A of 63 dB (20 μPa).

The reference measurements of all channels were performed using the ES method. The sweep duration was 2 s for all channels. The results of three representative channels are measured in Fig. 11. Fig. 11(a) shows a series of HIRs for channel 16 (elevation 5°), which represents the channel with the longest reverberation of the IR, $L_1 = 90$ ms. Fig. 11(b) shows a series of HIRs for channel 5 (elevation 10°), which represents the channel with the longest reverberation of the second-order HIR, $L_2 = 10$ ms. Fig. 11(c) shows the series of HIRs for channel 11 (elevation 70°). This channel showed the highest number of harmonics K for all channels. The threshold to identify HIRs was chosen at 70 dB below the peak amplitude. Comparing these three panels, it is evident that the parameters L_1 and L_2 did not vary noticeably.

The following parameters were used to configure the measurement with MESM: $L_1 = 100$ ms, $L_2 = 10$ ms, and $K = 5$. Applying these parameters to Eq. (15) results in $\eta = 3$ and $T' = 1.815$ s. This implies seven overlapped groups of three interleaved sweeps and one additional sweep. The excitation delays for each sweep were calculated according to Eq. (12). Then the system identification

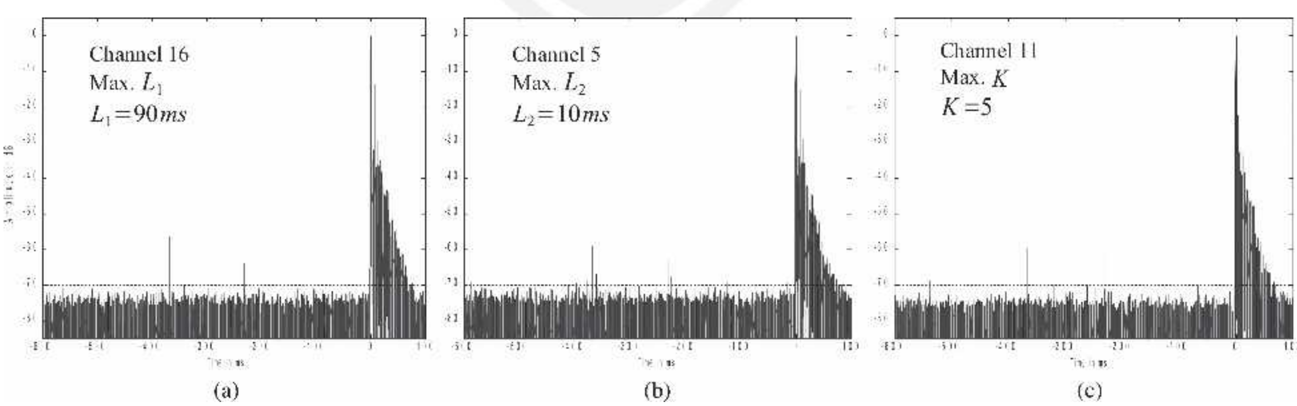

Fig. 11. Reference measurement using ES method; series of HIRs for different channels. (a) Channel 16 (elevation 5°), showing longest reverberation time of 90 ms in impulse response. (b) Channel 5 (elevation 10°), showing longest reverberation time of 10 ms in second-order HIR. (c) Channel 11 (elevation 70°), showing highest number of harmonics of $K = 5$ found in HIR series of all channels. Data were normalized and are presented in dB.

of all 22 channels was performed using the MESM. Fig. 12 shows the spectrogram of the recorded signal at the microphone. This signal was deconvolved using the inverse sweep to a series of HIRs, which is shown in Fig. 13. This figure shows a good separation of the IRs, whereas the higher order HIRs overlap and destroy the information about the nonlinear parts of the systems. The IRs representing the room IR for each channel were separated using appropriate windows.

As an example, the IR of channel 16 (elevation 5°) is shown in Fig. 14. For comparison the impulse response of the same channel measured with ES is shown in Fig. 15. This impulse response was obtained from the series of HIRs shown in Fig. 11 by windowing. Even though there should not be any difference between the two data, the IRs do not match each other perfectly because of the noise in the response data. To estimate the differences in a statistical way the relative root mean square error between the amplitude spectra within the interval (ω_1, ω_2) was investigated,

$$e_{\%} = \sqrt{\frac{\sum(|\boldsymbol{H}_{\text{MESM}}| - |\boldsymbol{H}_{\text{ES}}|)^2}{\sum|\boldsymbol{H}_{\text{ES}}|^2}}\,100\% \tag{19}$$

where $\boldsymbol{H}_{\text{MESM}}$ and $\boldsymbol{H}_{\text{ES}}$ are the transfer functions of a channel measured using MESM and ES, respectively. For all 22 channels the error was $e_{\%} = 0.6 \pm 0.06\%$. This is

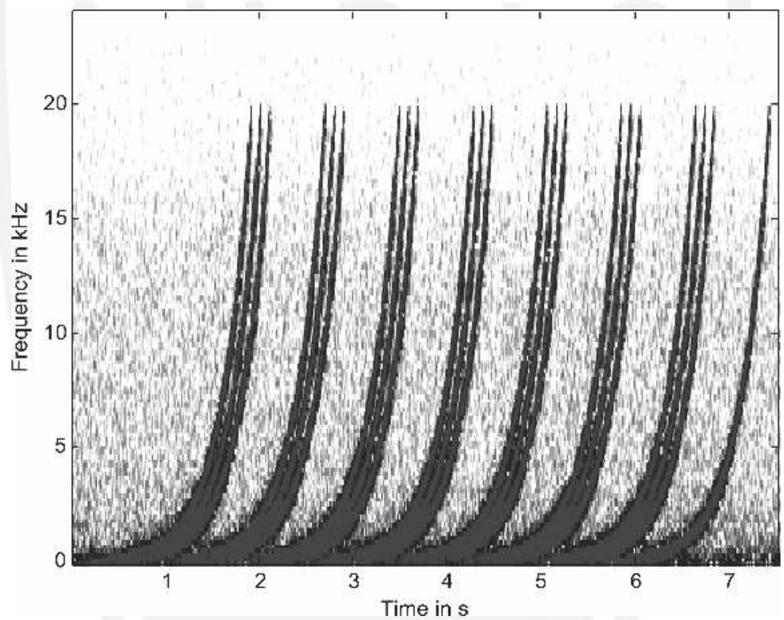

Fig. 12. Reference measurement using MESM. Recorded-signal spectrogram during excitation of 22 channels.

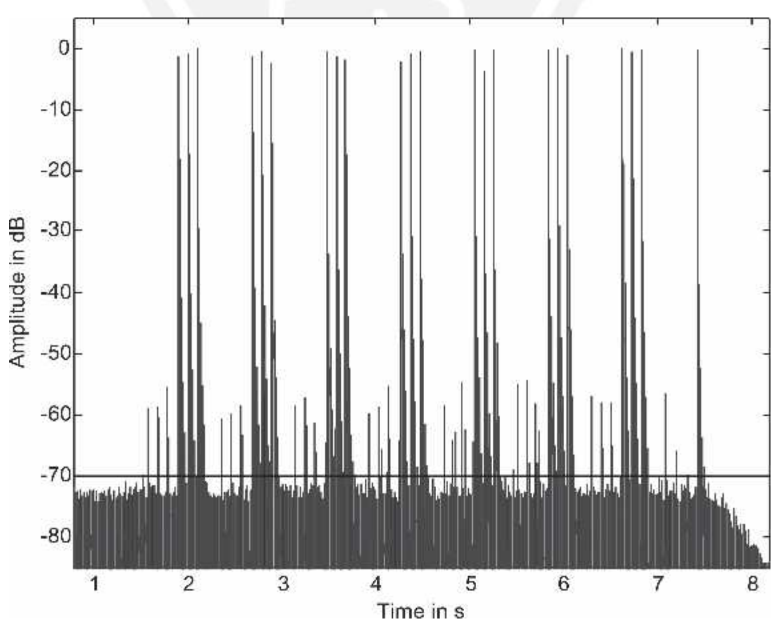

Fig. 13. Series of HIRs obtained from signal of Fig. 12. Data were normalized and are presented in dB.

within the range of the measurement repeatability[6] of 0.7 ± 0.7%. It actually shows that there is no significant difference between these two measurement methods.

The measurement duration of all 22 channels was $T_{\mathrm{MESM}} = 7.103$ s. Compared to the measurement duration for the unoptimized method ($T_{ES} = 35.2$ s for $T = 1.5$ s), a reduction by factor of five could be achieved.

The measurement of an HRTF set includes HRTFs for 1550 different spatial positions. The positions are distributed on the sphere with a constant spherical angle. The 22 channels of our system corresponded to elevations from −30° to +80° in steps of 5°. The resolution in the horizontal plane was 2.5°. On average 11 channels were measured for one horizontal position. The measurement duration of an HRTF set using the ES method would be 41 min. With MESM the measurement duration could be reduced to 10 min, which is only 24% of the duration using the unoptimized method.[7]

The HRTF measurement was configured to optimize the measurement duration. However, the MESM can also be configured to increase the SNR of a measurement. In such cases $\eta_{\mathrm{opt,SNR}}$ is calculated by solving Eq. (17). In this

[6]The repeatability of the system identification was investigated by calculating the relative error between 120 measurement repetitions.

[7]The time required to rotate the subject on the turntable was about 2 min, and it influences the measurement duration equally for both methods. Thus it is not included in the calculations.

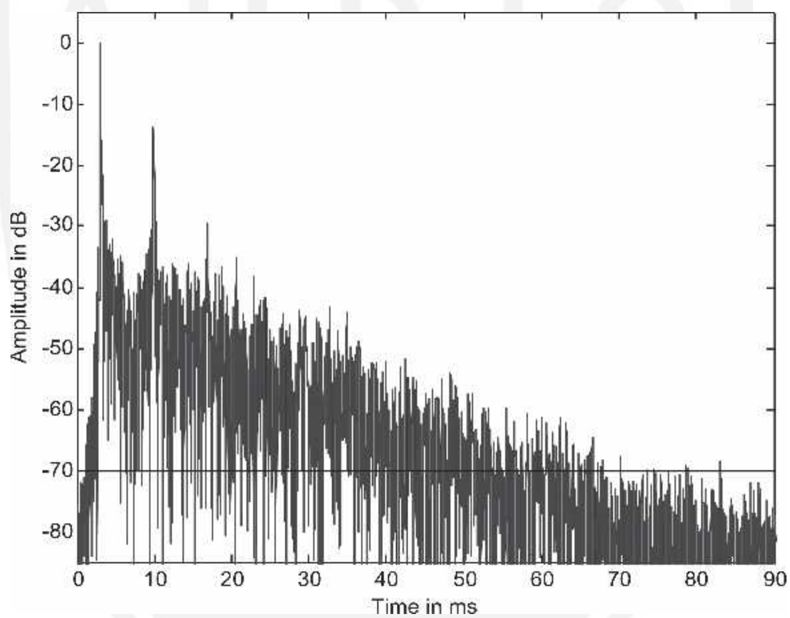

Fig. 14. Logarithmic and normalized room impulse response for channel 16 (elevation 5°) obtained with MESM.

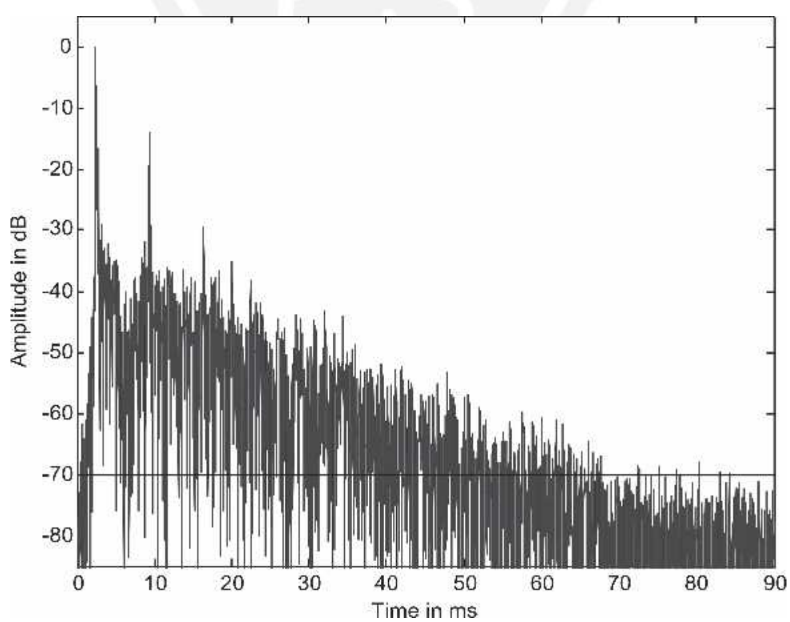

Fig. 15. Logarithmic and normalized room impulse response for channel 16 (elevation 5°) obtained with ES.

example this resulted in $\eta_{opt,SNR} = 22$ and $T' = 33$ s. The achieved SNR gain was $g = 13.4$ dB, although the measurement of 22 channels with MESM takes the same time as the measurement without optimization ($T_{MESM} = T_{ES} = 35.2$ s).

A comparison of the SNR gain and the improvement in measurement duration for interleaving different numbers of sweeps is shown in Tab 1.

5 CONCLUSIONS AND PERSPECTIVES

A new method for system identification, the multiple exponential sweep method (MESM), was introduced. MESM was developed for the system identification of multiple weakly nonlinear electroacoustic systems where linear transfer functions are of interest. MESM is based on the system identification using exponential sweeps. It shows the largest improvement in measurements where long excitation signals are required. Thus MESM is recommended for measurements under noisy conditions and/or in reverberant rooms. MESM can be optimized in two modes—to obtain the shortest measurement duration without reducing the SNR or to achieve as much SNR as possible without extending the measurement duration. MESM shows an excellent robustness against nonlinear distortions. The linear responses of the systems are measured without any artifacts of the nonlinear parts, resulting in the transfer functions of the measured systems. With respect to the question of sensitivity to the time variance, one may expect the same robustness to the time variance for the MESM as for the ES method. However, assuming the time variance to be a stochastic process in time, the artifact probability is lower for MESM due to the shorter measurement duration. Thus, on average, MESM is more robust than the ES method for weakly time-variant systems.

As an application of MESM, the measurement duration of the HRTF measurement was optimized. The measurement setup consisted of 22 loudspeakers, which allowed the HRTF measurement of 22 different elevations at once. The measurement duration for the 22 HRTFs could be reduced to 20% of the duration required for the unoptimized method. Measuring an HRTF set with 1550 positions, the measurement duration was reduced by a factor of four. Therefore this optimization mode is applicable for the fast identification of multiple systems such as HRTFs in a semianechoic sound chamber using ordinary multichannel equipment.

Furthermore the second optimization mode of MESM, achieving a higher SNR, was investigated based on the HRTF measurement equipment. Compared to the unoptimized method, the SNR could be increased by 13.4 dB. This optimization mode can be used for the identification of multiple systems, where the highest possible SNR is required and the measurement duration is not relevant.

In general MESM can be applied to different system identification tasks. For example, MESM can be applied for the measurement of the spatial characteristics of microphones, where excitation signals from different spatial positions are recorded by the test microphone [33]. A similar setup can be used for the measurement of room IRs for different excitation positions. An example for such a setup is the synthesis of sound fields in rooms [34], [35]. Also the in situ measurement of absorption and reflection coefficients requires excitations from different angles of incidence [36] and uses a similar setup.

Furthermore MESM can be analyzed in the context of linear system theory [37], where the MESM setup represents an application of a single-input multiple-output (SIMO) system [38], [39]. If the excitation signals are simultaneously recorded at different spatial positions, then the SIMO system is extended to a multiple-input multiple-output (MIMO) system. Actually our HRTF measurement setup is an example of a MIMO system, because the signals for the left and right ears are recorded simultaneously. From this point of view MESM can be applied to different system identification tasks of acoustic MIMO systems used for beamforming, source separation, or echo cancellation [34].

As an outlook, the system identification with exponential sweeps provides much room for improvements during the postprocessing. For example, further SNR improvements may be achieved by suppressing noise and extracting the relevant parts from the recorded signal using time-variant filters such as Gabor multipliers [40]. First heuristic tests showed some improvement from denoising in low-energy regions of measured spectra.

Table 1. Measurement duration T_{MESM} and SNR gain g for interleaving of η number of systems.*

η	T' (s)	T_{MESM} (s)	g (dB)
1	1.5†	12.162	0
2	1.5†	7.729	0
3	1.815	**7.103**	0.828
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
SNR gain optimization	33	35.2	**13.4**

* T' shows the sweep duration. The shortest measurement duration and the highest SNR gain are shown in bold. The measurement duration without optimization was $T_{ES} = 35.2$ s.

† According to Eq. (6) T' was lower in these cases, but it was set to 1.5 s to fulfill the SNR requirements.

6 ACKNOWLEDGMENT

This study was supported by the Austrian Academy of Sciences. The authors would like to thank Robert Höldrich for valuable suggestions on this work.

7 REFERENCES

[1] H. Møller, "Fundamentals of Binaural Technology," *Appl. Acoust.,* vol. 36, pp. 171–218 (1992).

[2] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of Human Subjects," *J. Audio Eng. Soc.,* vol. 43, pp. 300–321 (1995 May).

[3] F. L. Wightman and D. J. Kistler, "Headphone

Simulation of Free-Field Listening. I: Stimulus Synthesis," *J. Acoust. Soc. Am.,* vol. 85, pp. 858–867 (1989 Feb.).

[4] F. L. Wightman and D. J. Kistler, "Factors Affecting the Relative Salience of Sound Localization Cues," in *Binaural and Spatial Hearing in Real and Virtual Environments,* R. H. Gilkey and T. R. Anderson, Eds. (Lawrence Erlbaum, Mahwah, NJ, 1997).

[5] F. L. Wightman and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation," *J. Acoust. Soc. Am.,* vol. 85, pp. 868–878 (1989 Feb.).

[6] W. M. Hartmann and A. Wittenberg, "On the Externalization of Sound Images," *J. Acoust. Soc. Am.,* vol. 99, pp. 3678–3688 (1996 June).

[7] D. R. Begault and E. M. Wenzel, "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source," *J. Audio Eng. Soc.,* vol. 49, pp. 904–916 (2001 Oct.).

[8] A. W. Bronkhorst, "Localization of Real and Virtual Sound Sources," *J. Acoust. Soc. Am.,* vol. 98, pp. 2542–2553 (1995 Nov.).

[9] 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* (New Paltz, NY, 2001 Oct.) pp. 99–102.

[10] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization Using Nonindividualized Head-Related Transfer Functions," *J. Acoust. Soc. Am.,* vol. 94, pp. 111–123 (1993 July).

[11] J. C. Middlebrooks, "Virtual Localization Improved by Scaling Nonindividualized External-Ear Transfer Functions in Frequency," *J. Acoust. Soc. Am.,* vol. 106, pp. 1493–1510 (1999 Sept.).

[12] M. Morimoto and H. Aokata, "Localization Cues in the Upper Hemisphere," *J. Acoust. Soc. Jpn. (E),* vol. 5, pp. 165–173 (1984).

[13] J. C. Makous and J. C. Middlebrooks, "Two-Dimensional Sound Localization by Human Listeners," *J. Acoust. Soc. Am.,* vol. 87, pp. 2188–2200 (1990 May).

[14] W. G. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," *J. Acoust. Soc. Am.,* vol. 97, pp. 3907–3908 (1995 June).

[15] D. R. Begault, *3D Sound for Virtual Reality and Multimedia* (Academic Press, Cambridge, MA, 1994).

[16] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, "Interpolation Transfer Function for 3D-Sound Generation," *J. Audio Eng. Soc.,* vol. 52, pp. 915–930 (2004 Sept.).

[17] P. Minnaar, J. Plogsties, and F. Christensen, "Directional Resolution of Head-Related Transfer Functions Required in Binaural Synthesis," *J. Audio Eng. Soc.,* vol. 53, pp. 919–929 (2005 Oct.).

[18] E. H. A. Langendijk and A. W. Bronkhorst, "Fidelity of Three-Dimensional Sound Reproduction Using Virtual Auditory Display," *J. Acoust. Soc. Am.,* vol. 107, pp. 528–537 (2000 Jan.).

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

[20] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems* (Wiley, New York, 1980).

[21] 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 (1977 June).

[22] S. Golomb, *Shift Register Sequences* (Holden-Day, San Francisco, CA, 1967).

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

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

[25] D. D. Rife and C. Dunn, "Comments on Distortion Immunity of MLS-Derived Impulse Response Measurements," *J. Audio Eng. Soc.* (*Letters to the Editor*), vol. 42, pp. 490–497 (1994 June).

[26] J. Vanderkooy, "Aspects of MLS Measuring Systems," *J. Audio Eng. Soc.,* vol. 42, pp. 219–231 (1994 Apr.).

[27] M. Wright and J. Vanderkooy, "Comments on 'Aspects of MLS Measuring Systems'," *J. Audio Eng. Soc.* (*Letters to the Editor*), vol. 43, pp. 48–49 (1995 Jan./Feb.).

[28] J. S. Walker, *Fast Fourier Transforms* (CRC Press, Ann Arbor, MI, 1991).

[29] B. Zhou, D. M. Green, and J. C. Middlebrooks, "Characterization of External Ear Impulse Responses Using Golay Codes," *J. Acoust. Soc. Am.,* vol. 92, pp. 1169–1171 (1992 Aug.).

[30] P. Zahorik, "Limitations in Using Golay Codes for Head-Related Transfer Function Measurement," *J. Acoust. Soc. Am.,* vol. 107, pp. 1793–1796 (2000 Mar.).

[31] J. Vanderkooy, "Another Approach to Time-Delay Spectrometry," *J. Audio Eng. Soc.,* vol. 34, pp. 523–538 (1986 July/Aug.).

[32] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstract's*), vol. 48, p. 350 (2000 Apr.), preprint 5093.

[33] M. Dickreiter, *Handbuch der Tonstudiotechnik* (Saur, Munich, Germany, 1990).

[34] S. Yon, M. Tanter, and M. Fink, "Sound Focusing in Rooms: The Time-Reversal Approach," *J. Acoust. Soc. Am.,* vol. 113, pp. 1533–1543 (2003 Mar.).

[35] C. Falch, M. Noisternig, S. Warum, and R. Höldrich, "Room Simulation for Binaural Sound Reproduction Using Measured Spatiotemporal Impulse Responses," presented at the 6th Int. Conference on Digital Audio Effects, London, UK (2003 Sept.).

[36] E. Mommertz, "Angle-Dependent In Situ Measurements of Reflection Coefficients Using a Subtraction Technique," *Appl. Acoust.,* vol. 46, pp. 251–263 (1995).

[37] C-T. Chen, *Linear System Theory and Design* (Oxford University Press, New York, 1998).

[38] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing* (Springer, Berlin, 2006).

[39] H. Bessai, *MIMO Signals and Systems* (Springer, Berlin, 2005).

[40] P. Balazs, "Basic Definition and Properties of Bessel Multipliers," *J. Math. Anal. Appl.,* vol. 325, pp. 571–585 (2007).

THE AUTHORS

P. Majdak

P. Balazs

B. Laback

Piotr Majdak was born in Opole, Poland, in 1974. He studied audio engineering at the University of Technology and at the University of Music and Dramatic Arts, both in Graz, Austria, and received an M.Sc. degree in 2002. Currently he is finishing his Ph.D. degree at the Institute for Electronic Music and Acoustics, Graz, on the effects of fine-structure interaural time differences on lateralization in electrical hearing.

During his studies he developed software and hardware for atmosphere controlling devices. Since 2002 he has been working at the Acoustics Research Institute, Austrian Academy of Sciences, and is involved in projects including binaural signal processing, localization of sounds, and cochlear implants.

Mr. Majdak is a member of the Audio Engineering Society and the Acoustical Society of America.

●

Peter Balazs was born in Tulln, Austria, on December 11, 1970. He received M.S. (2001) and Ph.D. (2005) degrees in mathematics, both with distinction, from the University of Vienna, Vienna, Austria.

He joined the Acoustics Research Institute of the Austrian Academy of Science in 1999. He was a fellow of the HASSIP EU network cooperating with NuHAG, faculty of mathematics, University of Vienna.

Dr. Balazs is the author of four journal and 11 conference papers. He is interested in time–frequency analysis, Gabor analysis, numerics, frame theory, signal processing, acoustics, psychoacoustics, system identification, and speech coding. He is a member of the Audio Engineering Society.

●

Bernhard Laback was born in Vienna, Austria, on May 27, 1970. He received a Ph.D. degree from the University of Vienna in 1999, after studying natural-sciences-oriented musicology and sound engineering.

He joined the Acoustics Research Institute in the same year, starting psychoacoustic research on cochlear implants. He built up and currently leads the Experimental Audiology Group, focusing on binaural as well as monaural cochlear implant research. His main interests are the sensitivity to spatial auditory cues, but he also worked on masking and channel interaction phenomena in cochlear implant and normal hearing listeners. Together with his colleagues, he published several peer-reviewed research articles and has been an invited speaker at several conferences. He established national and international cooperation with national and international basic research institutions, ENT departments, and cochlear implant manufacturers.

Dr. Laback is a member of the Association for Research in Otolaryngology.