forum acusticum 2026
8-12 September · Graz, Austria

NONLINEAR HARMONIC PREDICTION AND COMPENSATION IN HRTF MEASUREMENTS

Payman Azaripasand[1], Bernhard U. Seeber[1]

[1] *Audio Information Processing, Technical University of Munich, Germany*

Abstract

The exponential sine sweep (ESS) technique became the common approach for measuring head-related transfer functions (HRTFs) since it allows separating the loudspeaker's distortion from the HRTF by temporal windowing. This, however, fails with overlapping excitation signals. This work assesses the signal-to-noise ratio (SNR) benefit from nonlinear harmonic prediction and subtraction. The measured response is modeled as a superposition of a linear response, nonlinear harmonic distortions, and additive noise. Each harmonic distortion component is represented using a finite impulse response kernel. The kernel is estimated from a reference measurement. For evaluation, broadband and Bark-band limited SNR is computed. Preliminary results for frontal HRTF measurements show broadband SNR values of 30–45 dB depending on the loudspeaker direction, with the proposed harmonic subtraction improving these values by up to about 3 dB for the most affected (most elevated) channel. In perceptually relevant mid-to-high frequency regions, where spectral notches make the HRTF sensitive to measurement noise, mean SNR gains of 1–2 dB are observed. The results demonstrate that the proposed method has potential to reduce non-linear distortion and provide cleaner impulse responses and more accurate noise estimates in overlapping ESS-based HRTF measurements.

Keywords: *Head-Related Transfer Functions, Harmonic Distortions, SNR-estimation*

1. Introduction

Head-related transfer functions (HRTFs) describe the acoustic filtering effect of the human head, torso, and pinnae, and are essential for spatial audio reproduction and binaural rendering. Accurate measurement of HRTFs requires a high signal-to-noise ratio (SNR) across a wide frequency range in order to correctly capture phase and amplitude information, particularly around spectral notches. Several measurement techniques have been proposed for obtaining impulse responses in acoustic systems. Among them, exponential sine sweep (ESS) excitation has become a widely used approach due to its high dynamic range and its ability to separate nonlinear distortion components after deconvolution. In ESS measurements, harmonic distortions generated by the playback system appear at predictable temporal offsets relative to the linear impulse response, allowing them to be analyzed separately [1, 2]. When measuring HRTFs, measurement time is often reduced by using time-optimized playback strategies by overlapping measurement sequences played from different loudspeakers. While improving measurement time efficiency, these approaches introduce issues with SNR and additional complexity because the linear impulse response and nonlinear distortion components originating from different excitations may overlap in time. As a result, these deterministic signal components can no longer be readily isolated and thus alter the measured impulse response and the residual measurement noise floor. Conventional SNR estimation methods typically rely on time gating of the impulse response tail, on comparing playback and recorded energy, or on the analysis of crest factors in the impulse response signal under the assumption of uncorrelated noise. Furthermore, the permissible overlap of successive sweeps has traditionally been limited to ensure that harmonic components remain sufficiently separated in time [3].

Traditionally, non-linearities in system identification have been modeled using the Volterra or Hammerstein system representations, where the output is described as a sum of responses corresponding to different nonlinear orders of the input signal [4]. While theoretically robust, the computational complexity of full Volterra kernel estimation [5] often limits their applicability, yet it is shown to be efficient for modelling distortions for low frequency operating drivers such

Corresponding author: *p.azaripasand@tum.de.*

as subwoofers [5, 6]. As a computationally lighter alternative to full nonlinear system identification, researchers investigated the temporal separability of harmonic distortions inherent in ESS measurements and their application to system identification [1, 2]. However, this line of work has focused on isolating the distortion components themselves in situations with low temporal overlap rather than on quantifying measurement quality particularly when distortion components overlap; consequently, for these overlapping scenarios a robust quality evaluation metric, such as an SNR estimation, would be beneficial.

In this work, we introduce a deterministic component-separation framework for ESS-based HRTF measurements that enables reliable SNR estimation even when impulse responses overlap in time. Also, the same decomposition yields a corrected impulse response with neighboring-sweep and harmonic contamination reduced – a benefit that extends to cases where late reflections might coincide with harmonic distortion components. Rather than identifying a nonlinear system model, the proposed method exploits the deterministic temporal structure of ESS measurements to model the measured head-related impulse response as a superposition of time-shifted harmonic templates derived from a reference measurement. Using the a priori known ESS timing, overlapping linear and harmonic contributions from neighboring sweeps are progressively identified and removed together with the harmonic distortion of the target response. Harmonic components predicted from corresponding references are subtracted to obtain a cleaned HRTF estimate and the residual, enabling wideband and Bark-band SNR estimation directly from the impulse response without dedicated noise recordings or non-overlapping sweeps.

2. Methods

2.1. Exponential Sine Sweep measurement

Head-related transfer functions were measured using an ESS excitation, defined as [7]:

$$s(t) = \sin\left(\frac{2\pi f_1 T}{\ln(f_2/f_1)}\left(e^{t\ln(f_2/f_1)/T} - 1\right)\right) \quad (1)$$

where f_1 and f_2 denote the start and end frequencies of the sweep respectively, and T is the sweep duration. The logarithmic frequency progression allows harmonic distortions generated by the playback chain to appear at distinct temporal positions after deconvolution which is a specific property of ESS signals.

$$h(t) = y(t) * S^{-1}(t) \quad (2)$$

To obtain the impulse response, the recorded signal $y(t)$ is convolved with the inverse sweep $S^{-1}(t)$.

2.2. Temporal structure of harmonic distortions

A key property of ESS measurements is that nonlinear distortions are separated in time after linear deconvolution. The impulse response corresponding to the k-th harmonic distortion appears earlier than the linear response at a time offset given by

$$\Delta t_k = -\frac{T}{\ln(f_2/f_1)}\ln(k) \quad (3)$$

where $k = 1$ corresponds to the linear impulse response and $k > 1$ to higher-order harmonic distortions, where the discrete-time equivalent is Δn_k.

2.3. Impulse response model

The measured impulse response can be modeled as a superposition of the linear response, harmonic distortions, and residual noise as

$$h[n] = \sum_{k=1}^{K} c_k\, h_0[n - \Delta n_k] + \epsilon[n] \quad (4)$$

where

- $h_0[n]$ is the linear impulse response,
- c_k are scaling coefficients representing the contribution of each harmonic component,
- Δn_k are the discrete time known harmonic delays,
- K is the number of modeled harmonic components,
- $\epsilon[n]$ representing a residual noise.

2.4. Overlapping sweep interference

In order to reduce the total measurement time, modern HRTF acquisition systems often employ overlapping sweep playback, where multiple exponential sine sweeps are emitted sequentially with partial temporal overlap. While this strategy improves measurement efficiency, it introduces additional complexity in the impulse response domain after deconvolution.

Let the playback times of consecutive, identical sweeps be denoted by n_i. After deconvolution, each excitation produces a set of impulse responses consisting of the linear response and its harmonic distortions. For sweep i the linear impulse response appears at time sample n_i while its harmonic distortion responses occur at time offsets Δn_k determined by the sweep parameters introduced above. When overlapping playback is used, the impulse responses of neighboring sweeps may partially overlap in time. Consequently, as illustrated in Fig. 1, the impulse response segment associated with a particular loudspeaker channel may contain contributions from

- the linear impulse response of the investigated sweep,
- harmonic distortions of the same sweep,
- harmonic distortions of neighboring sweeps,
- linear responses of adjacent sweeps.

And hence the impulse response model from Eq. (4) will be

$$h[n] = \sum_i \sum_k c_{i,k}\, h_0[n - (n_i - \Delta n_k)] + \epsilon[n] \quad (5)$$

n_i is the sample index of the i-th sweep playback. Because the playback schedule and harmonic are known, the temporal positions of all overlapping components can be predicted. The impulse response segment of the investigated loudspeaker is identified from the known playback timing. A window centered around the main impulse response is then selected with a duration matching the expected full impulse response length for that loudspeaker channel after deconvolution. Within this interval, neighboring-sweep contri-

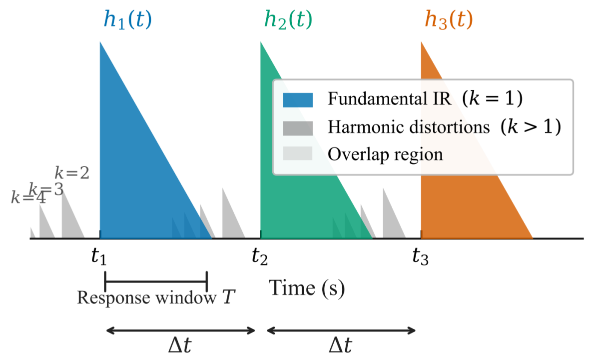

Figure 1: Sketch of overlapping impulse responses along with their respective harmonic distortions on their left side as the result of linear deconvolution.

butions are treated as interference. To reduce the IR contamination and avoid bias in the SNR estimate, the proposed method explicitly includes all interfering components inside the analysis window. Impulse responses and harmonic responses belonging to adjacent sweeps are added as separate components in the decomposition model and their amplitudes are estimated during the fitting process. After the fitting step, contributions of these adjacent-sweep components are then subtracted from the impulse response under investigation. The remaining residual signal after removing all the extra modeled components is then interpreted as measurement noise.

2.5. Neighbor-overlap identification

When overlapping ESS measurements are employed, impulse responses from adjacent playback channels may fall inside the deconvolution interval of another channel. To identify these interactions, a neighboring-overlap matrix was introduced and computed. Hence, for each channel i, the valid interval in the deconvolved impulse response was defined as:

$$\Omega_i = [p_i - (N_{sweep} - 1),\, p_i + (N_{interval} - 1)] \quad (6)$$

where

- p_i denotes the peak location (detected for impulse responses and calculated for nonlinear harmonics),
- N_{sweep} is the excitation length,
- $N_{interval}$ is the sweep repetition interval.

Each extracted impulse response occupies

$$\Gamma_j = [p_j - L_1,\, p_j + L_2]\,, \quad (7)$$

where $L_1 + L_2$ equals the total samples considered for the impulse response or harmonics extraction. An overlap exists whenever:

$$\Gamma_j \cap \Omega_i \neq \emptyset, \quad (8)$$

or equivalently,

$$\Gamma_{j,left} \leq \Omega_{i,right} \quad \text{and} \quad \Gamma_{j,right} \geq \Omega_{i,left} \quad (9)$$

This test was repeated independently for

- the linear impulse response,
- every harmonic order.

Then we will have for linear responses N_{IR} and $N_H^{(k)}$ for harmonic order k. These matrices identify only physically overlapping components, preventing wrong subtraction, and providing the basis for the progressive, order by order neighbor removal.

2.6. Regularized harmonic fitting and subtraction

To estimate the contribution of each harmonic component, the measured impulse response is modeled as the superposition of deterministic components and an additive residual noise term

$$\mathbf{h} = \mathbf{A}\mathbf{c} + \epsilon \quad (10)$$

where the columns of the matrix $\mathbf{A}$ contain time-shifted harmonic templates corresponding to the linear response and the harmonic distortion components and the coefficient vector $\mathbf{c}$, representing the contribution of the direct impulse response, its harmonic distortions, and the corresponding neighboring-sweep components, is estimated through an optimization procedure and the resulting scaled templates are subtracted from the measured impulse response. This is applied progressively, in order: the direct impulse response, neighboring impulse responses, predicted harmonic distortion components, and neighboring harmonic components. The residual signal is obtained as:

$$\begin{aligned} h_{noise}[n] &= h[n] - h_{main}[n] - h_{harm}[n] \\ &\quad - h_{main,neighbors}[n] - h_{harm,neighbors}[n] \quad (11) \end{aligned}$$

which is interpreted as the measurement noise and forms the basis for the SNR estimation.

2.7. SNR estimation

Band-limited SNR values are computed by integrating the power spectral density of the signal and noise components over predefined frequency bands, wide band as well as Bark critical bands [8].

2.8. Measurement setup and data acquisition

The data analyzed in this study were obtained from a set of head-related transfer function (HRTF) measurements conducted in the anechoic chamber of the Audio Information Processing Group at the Technical University of Munich. Additional details on the measurement setup are provided in [9]. A human subject was positioned at the center of a circular loudspeaker array mounted on a vertical rotating arc, where the arc placed in the median plane. The playback and recording system consisted of an RME MADIface USB audio interface (RME GmbH, Haimhausen, Germany) connected via MADI to an RME 12Mic preamplifier (RME GmbH) for the microphone inputs, and to a 32-channel MA32D2 digital audio amplifier (Innosonix GmbH, Ahorn, Germany) for loudspeaker playback. The loudspeaker array comprised 32 VISATON FRS8M full-range drivers (Visaton GmbH & Co. KG, Haan, Germany) mounted in custom-built enclosures. Prior to the measurements, all loudspeakers were equalized using a minimum-phase inverse filter derived from individual free-field measurements, ensuring a flat magnitude response within the operating bandwidth. In-ear microphones based on the Sennheiser KE 4-211-2 capsule (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) were used to record the ear signals. According to the specifications, the microphone self-noise corresponds to 27 dB(A), which defines a part of the intrinsic noise floor of the recording system. The loudspeakers were positioned on a vertical arc around the subject at a distance of 1.35 m from the center of the head, covering a range of elevations in the median plane. The excitation signals using ESS with duration of 500 ms and an overlap ratio of 50 percent and covering a frequency range of 150–20000 Hz were played at approximately 75 dB SPL calibrated at 1 kHz at the head center. As for reference (no HRTF), the measurement was conducted under identical setup conditions using non-overlapping 5 seconds long excitations. The recordings were performed at a sampling frequency of 44.1 kHz and were band-pass filtered using a zero-phase FFT filter applied in the same frequency range as the sweep signal, ensuring removal of out-of-band noise without introducing phase distortions. The low-frequency background noise in the anechoic chamber was negligible compared to the playback level and was further suppressed during the band-pass filtering.

3. Results

3.1. Impulse response calculation and harmonic structure

The impulse responses of the 32 loudspeakers positioned on the median plane were obtained using linear ESS deconvolution, as in Eq. (2), under the overlapping conditions described in the previous section. Owing to the known playback timing, the deconvolved impulse responses are clearly separated and easily identified. These measurements represent an actual HRTF measurement condition for the proposed method: within the analysis window of the investigated loudspeaker only its linear impulse response and the corresponding harmonic components are present. All contributions from other loudspeakers are treated as external components and are computed separately for their respective channels. Depending on the playback schedule and the number of modeled harmonics, these external components may or may not fall into the analysis window depending on the overlap ratio. Figure 2 shows the reconstructed impulse responses for the left ear at frontal direction. The impulse response consists of the linear response together with harmonic distortion components generated by the playback chain for three loudspeaker channels. Using the known analytical relationship between sweep parameters and harmonic timing, the expected temporal positions of harmonic distortions up to the 10th order were calculated. Distortion components occurring at these positions were then identified and included in the component-separation model described in Section 2.
In the present measurements, a peak factor [10] of approximately 70 dB can be observed in Fig. 2. The tail of the impulse response, including an early reflection lying outside the 512-sample impulse response window, also contributes to the residual noise in this figure. It should be noted that these measurements were conducted using overlapping sweeps with relatively short sweep durations and without interleaving strategies. Despite the reduced measurement time, the impulse responses still exhibit clearly identifiable harmonic structures, enabling the proposed method.

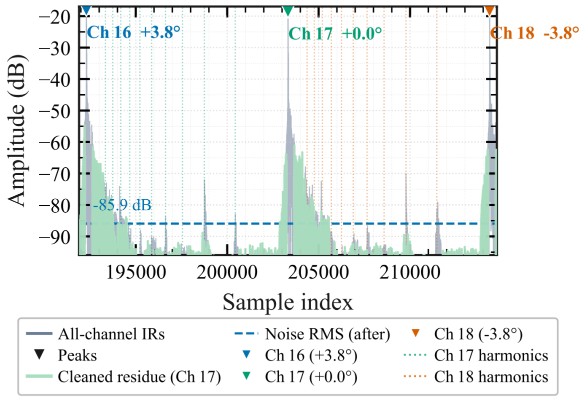

Figure 2: Calculated Head-Related Impulse Responses obtained from a median-plane measurement ($\varphi = 0°$) using a 500 ms excitation signal with 50% overlap. The impulse response of the 0.0° elevation loudspeaker channel is visible within the plotted time range, together with components from neighboring loudspeaker channel impulse responses or harmonics, as well as the residual noise (post-subtraction) of the corresponding channel.

Table 1: Relative energy contributions of signal, distortion, and residual noise across measurement conditions. Values are reported as mean ± standard deviation across 32 loudspeaker channels.

Condition	Dist./Main (%)	Noise/Main (%)	SNR (dB)
Median – L ear	0.06 ± 0.17	1.7 ± 4.5	40.9 ± 4.1
Median – R ear	0.04 ± 0.13	1.2 ± 2.8	41.8 ± 3.4
R side – L ear	0.44 ± 0.29	11.4 ± 9.5	31.3 ± 4.8
R side – R ear	0.06 ± 0.24	0.7 ± 1.1	43.6 ± 2.8

3.2. Separation of signal, distortion and noise

The proposed method separates the measured impulse response into three components: the linear impulse response, harmonic distortions, and residual noise. The main impulse response component represents the linear acoustic transfer function. Harmonic distortions appear at the expected ESS time offsets, while the remaining signal after removing these components as in Eq. (11), represents the residual noise. The energy of each component can then be evaluated independently, allowing the SNR to be computed directly from the impulse response domain. As can be seen in Table 1, the linear impulse response contains the dominant portion of the signal energy across all measurement conditions. Across the 32 loudspeaker channels, the harmonic distortion energy remains consistently below 0.5% of the main impulse response energy, indicating that the loudspeakers used in this setup exhibit low nonlinear distortion at the playback level. The residual noise component accounts for a small share of the signal energy in frontal and ipsilateral directions, while higher values of up to 11% are observed in contralateral directions due to acoustic shadowing effects and possible relatively more difficult fitting of reference distortions. These variations correspond closely to the observed directional differences in signal-to-noise ratio and reflect the underlying acoustic propagation conditions. Overall, the results demonstrate that the proposed method reliably isolates the dominant linear impulse response component while maintaining low distortion leakage and meaningful residual noise levels, enabling a consistent energy-based interpretation of measurement quality. This approach differs from conventional gating-based methods, where the separation between signal and noise relies on manually defined time windows.

3.3. SNR analysis

Prior to the measurements, all excitation signals were equalized by convolving them with the loudspeaker-specific inverse filters so that variations in SNR reflected the acoustic propagation and measurement conditions rather than differences in loudspeaker frequency response. The wideband SNR values obtained using the proposed method vary across loudspeaker directions due to differences in acoustic propagation paths and head-related filtering. For the frontal measurement (0° azimuth), the left ear exhibits wideband SNRs ranging from approximately 25 dB at the extreme upper and lower loudspeakers to approximately 45 dB at the center of the array. The SNR remains consistently above 30 dB for the majority of the loudspeakers (except the lower most channel) and the proposed subtraction method improves the SNRs by up to about 3 dB for extreme channels where overlapping deterministic components are most pronounced. Due to having low distortion loudspeakers, the absolute SNR enhancement introduced by the proposed method is generally below 1 dB for most channels; SNR improvements exceeding 2–3 dB are also observed for the extreme loudspeakers. The higher SNR values observed for directions close to the horizontal plane are expected as there are strong direct sound paths to the ear microphones. In contrast, bottom and elevated directions often involve stronger diffraction and shadowing effects caused by the body, head and pinna, resulting in attenuation and therefore lower SNR values. It is important to restate that these SNR values represent the ratio between the energy of the linear impulse response and the residual signal after progressively removing harmonic distortion components, rather than a direct comparison between playback level and environmental noise. Nevertheless, the linear model provides stable separation for all directions with clearly resolved harmonic timings. Consequently, the obtained SNR values directly quantify the effective quality of the recovered impulse response. To further investigate the frequency dependent measurement quality, the SNR was computed in Bark critical bands. The first Bark band was excluded since it lies below the effective frequency range of the loudspeaker equalization. For the frontal direction, Fig. 3 shows the left ear's spectral behavior: the Bark-band mean SNR increases from approximately 30 dB in the lowest bands to values exceeding 45 dB between approximately 4 and 8 kHz, before gradually decreasing toward 40 dB at the highest frequencies.

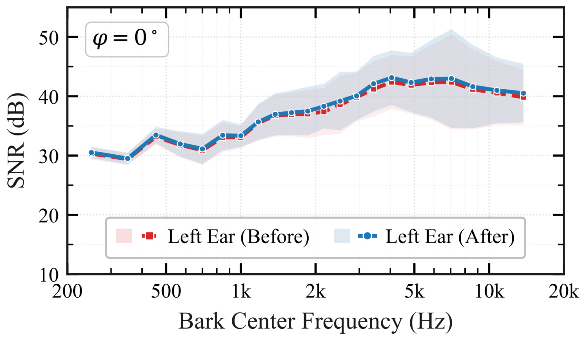

Figure 3: Bark-band SNRs calculated across 32 elevational loudspeakers on the median plane ($\varphi = 0°$) before and after the harmonic subtraction. Values were calculated from an HRTF measurement of a human subject and are analyzed and plotted per critical band within the measurement range.

Across nearly all Bark bands, harmonic component removal produces a consistent SNR increase. The improvement is relatively modest in the mid-frequency region, where the measurements are dominated by the linear impulse response, but becomes more noticeable at higher frequencies where harmonic distortion components contribute more strongly to the residual signal. Overall, the proposed decomposition method provides a consistent SNR improvement across the audible spectrum while preserving the expected acoustic characteristics of HRTF measurements. The largest absolute SNR values occur within the frequency region between approximately 3 and 8 kHz, where human auditory sensitivity is greatest.

4. Discussion

This work introduced a deterministic harmonic prediction and subtraction method for ESS-based HRTF measurements that jointly models the linear response, harmonic distortions, and neighboring-sweep interference. The proposed method offers three main advantages over conventional ESS approaches. First, the proposed method defines IR quality on the basis of signal energy rather than peak amplitude. Conventional approaches such as the Multiple Exponential Sweep Method (MESM) [10] do not compute the SNR analytically but infer it from the deconvolved impulse response by comparing the peak amplitude of the linear impulse response to the noise floor. The noise floor is typically estimated in regions where harmonic contributions fall below an identifiable threshold and become indistinguishable from noise. In MESM, this threshold is commonly set to approximately 70 dB below the main impulse response peak. Using this criterion, SNRs on the order of 80 dB have been reported under optimized measurement conditions, e.g. with sweep duration of approximately 1.5 s [3]. Such peak-based metrics can be misleading in HRTF measurements, since diffraction and multipath propagation around the head and pinna reduce the peak amplitude of the impulse response even though substantial signal energy, spread out in time, may still be present. By contrast, the proposed method explicitly separates the linear, harmonic, and residual components and defines SNR in terms of the total energy of the components, providing a more physically meaningful and less sampling, phase, or window-sensitive measure of measurement quality. Second, the proposed method also performs deterministic quantification of harmonic distortion. Because the temporal positions of harmonic distortion components after ESS deconvolution are known a priori, they can be identified and subtracted directly rather than inferred statistically, as in Volterra-based approaches. As shown in Section 3.2, the separated distortion energy remains low across all loudspeaker directions, confirming that the loudspeakers used in this setup exhibit low nonlinear distortion. Estimating the amplitude of each deterministic harmonic component before computing the residual reduces the risk of distortion energy being incorrectly attributed to the noise term, yielding a more consistent signal, distortion, and residual noise separation than simple time-windowing methods and improving the accuracy of the resulting SNR estimates. Third, the proposed method further remains applicable to time-optimized measurements employing overlapping ESS excitations. Because the temporal positions of both harmonic distortions and neighboring sweep responses are known a priori, interfering components originating from neighboring sweeps are progressively identified – via the neighbor-overlap matrix of Section 2.5 – and explicitly modeled as deterministic signal contributions rather than being implicitly absorbed into the noise estimate. These benefits should be read alongside some limitations. The loudspeakers used in this study exhibited low nonlinear distortion overall; systems with stronger nonlinearity may benefit more. The larger residual noise observed on the contralateral side also suggests that reference-kernel fitting might become less reliable under strong acoustic shadowing, which can be a limitation of the fitting phase rather than of the SNR definition itself. Finally, these results are based on a single subject and on the median plane; broader validation across subjects and directions is needed to confirm generality.

5. Conclusion

This work presented a method for obtaining corrected head-related impulse responses and consistent SNR estimates in ESS-based HRTF measurements. The method exploits the known temporal structure of ESS measurements to separate the linear response, harmonic distortion, and residual components. Based on this separation, an energy-consistent SNR definition is derived directly from the reconstructed signal and noise components, enabling a physically meaningful interpretation of measurement quality. The approach is particularly suited for HRTF acquisition systems using overlapping ESS excitations, where conventional time-gating or peak-based SNR metrics become unreliable. Besides, it provides non-underestimated SNRs, the decomposition yields corrected impulse responses, making it applicable to time-optimized measurements. The presented framework provides a practical, reproducible method for quantifying measurement quality in ESS-based HRTF acquisition, and the underlying decomposition concept may also be applicable to other acoustic measurement scenarios using swept-sine excitation.

6. Acknowledgments

Author PA was supported by the German Academic Exchange Service (DAAD), Program ID: 57552340, and the hardware setup was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330, Project C5.

References

[1] A. Novák, L. Simon, and P. Lotton, "Synchronized swept-sine: Theory, application, and implementation," *J. Audio Eng. Soc.*, vol. 63, no. 10, pp. 786–798, 2015. DOI: 10.17743/jaes.2015.0071.

[2] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, 2001.

[3] P. Dietrich, B. Masiero, and M. Vorländer, "On the optimization of the multiple exponential sweep method," *J. Audio Eng. Soc.*, vol. 61, no. 3, pp. 113–124, 2013.

[4] A. Novák, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 8, pp. 2220–2229, 2010. DOI: 10.1109/TIM.2009.2031836.

[5] W. J. Klippel, "Loudspeaker nonlinearities—causes, parameters, symptoms," *J. Audio Eng. Soc.*, vol. 54, no. 10, pp. 907–939, 2006.

[6] K. Lashkari and N. Naka, "Loudspeaker linearization using perceptual distortion measures," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 1, 2003, pp. 745–749. DOI: 10.1109/ACSSC.2003.1292014.

[7] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. 122nd Audio Eng. Soc. Conv.*, Vienna, Austria, 2007, Paper 7121.

[8] E. Zwicker, "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," *J. Acoust. Soc. Am.*, vol. 33, no. 2, p. 248, 1961. DOI: 10.1121/1.1908630.

[9] P. Azaripasand and B. U. Seeber, "HRTF measurement system with rotating loudspeakers," in *Fortschritte der Akustik – DAGA '25*, Deutsche Gesellschaft für Akustik e.V. (DEGA), 2025, pp. 221–224. DOI: 10.71568/dasdaga2025.402.

[10] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *J. Audio Eng. Soc.*, vol. 55, no. 7/8, pp. 623–637, 2007.