CHARACTERISATION AND MODELLING OF NON-LINEAR LOUDSPEAKERS

Leela K. Gudupudi[1]*, Christophe Beaugeant*[2] *and Nicholas Evans*[1]

[1]EURECOM, Sophia-Antipolis, France
`lastname@eurecom.fr`
[2]INTEL Mobile Communications, Sophia-Antipolis, France
`firstname.lastname@intel.com`

ABSTRACT

Approaches to non-linear acoustic echo cancellation for mobile devices all require a reliable, non-linear model of the loudspeaker. This paper investigates a recent approach to non-linear system identification. The loudspeaker is characterised using specially-crafted test signals, in this case simple exponential sine-sweep inputs, which facilitate the derivation of a polynomial Hammerstein model. Model performance is then assessed using the same artificial test input signal and then using real-speech. Model performance is assessed objectively using the mean cepstral distance between real loudspeaker outputs and that estimated using the model. Results show the potential of the approach but also the challenge in estimating reliable non-linear models which accurately predict the response to complex real-speech inputs.

***Index Terms*—** Nonlinearities, system identification, echo cancellation, loudspeaker modeling, exponential sine-sweep

1. INTRODUCTION

Non-linear acoustic echo cancellation (NAEC) has attracted growing attention over recent years [1–4]. This is perhaps due to the increased use of miniature loudspeakers in mobile devices; unfortunately, small loudspeakers tend to introduce non-linear distortion which often degrades the performance of acoustic echo cancellation algorithms [5–9].

Approaches to NAEC depend fundamentally upon a non-linear model of the loudspeaker. Several models have been reported in the literature [10–15]. The most popular are based on Volterra series [16] given by:

$$x_{out}(n) = \sum_{p=1}^{N} \sum_{i_1=0}^{M-1} \cdots \sum_{i_p=0}^{M-1} h_p(i_1, \cdots, i_p) x(n-i_1) \cdots x(n-i_p) \tag{1}$$

where $x(n)$ and $x_{out}(n)$ are respectively the input and output of an N^{th} order non-linear system with memory length M. Non-linearity is characterised via the set of N multidimensional Volterra Kernels $h_p(i_1, \cdots, i_p)$.

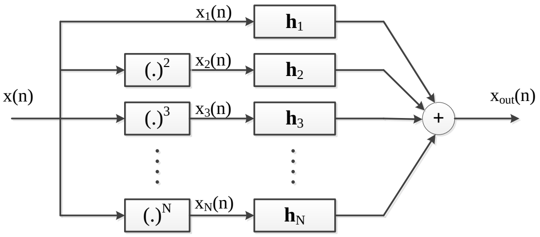

Fig. 1: The generalised polynomial Hammerstein model.

The Volterra series in Eq. 1 typically requires a large number of parameters to model practical non-linearity and is rarely used on account of impractical computational complexity [17]. However, the principal causes of nonlinearities in the mobile phones such as the clipping amplifier, nonlinear suspension and nonuniform flux density (in a loudspeaker) can be approximated as simplified memoryless models [18]. The popular power filter solution [2, 8] is given by:

$$x_{out}(n) = \sum_{p=1}^{N} \alpha_p x^p(n) \tag{2}$$

where α_1 is the loudspeaker gain and where α_p for $p > 1$ is the gain of the p^{th} order non-linear component. While the reduced complexity of the power filter model supports practical implementations, the reduced flexibility generally results in less accurate non-linear modelling, as shown in our previous work [19].

Alternative models with a more favourable compromise between computational demands and flexibility have thus been investigated. This paper reports our work using the polynomial Hammerstein model [20–22] illustrated in Fig. 1 and given by:

$$x_{out}(n) = \sum_{p=1}^{N} \sum_{i=0}^{L-1} x^p(n-i) h_p(i) \tag{3}$$

The linear filters $\mathbf{h}_p = [h_p(0), h_p(1), \cdots, h_p(L-1)]^T$ for $p \geq 1$ in Eq. 3 correspond to the diagonal Volterra Kernels in Eq. 1 [19,22] and can be readily identified according to procedures reported in [21–23]. This paper reports a comparison of real mobile device loudspeaker outputs measured in a controlled environment to those estimated by a polynomial

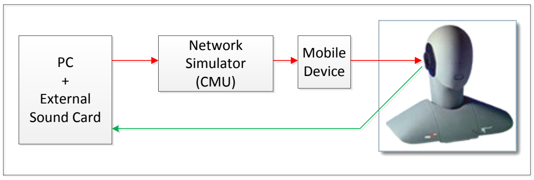

Fig. 2: Experimental setup to measure loudspeaker outputs.

Hammerstein model. The work involves three different mobile devices and reports the first investigation of model accuracy as a function of the key parameters, namely the number of filter taps L and the order of non-linearities N.

The reminder of this paper is organized as follows. Section 2 describes how model parameters are learned from loudspeaker test signals and how the model is applied in practice. A comparison of real loudspeaker signals to model estimates is presented in Section 3. Conclusions and perspectives are presented in Section 4.

2. NON-LINEAR LOUDSPEAKER MODELLING

This work aims to derive a generalised, non-linear loudspeaker model. We outline the process used to measure real loudspeaker outputs and to isolate their characteristics from those of recording effects. We then show how the model parameters are estimated from practical measurements and finally how the model is applied in practice.

2.1. Measurement

The experimental setup used for loudspeaker characterisation is illustrated in Fig. 2. A mobile device is placed before a head and torso mannequin at a distance of 32cm. The device is configured to operate in hands-free mode and at maximum volume for which non-linear distortion is assured. A PC is used to store and record all audio data sent to, or received from a mobile device via a high-quality external sound card and a network simulator [24]. Some additional non-intrusive tests confirmed that the nonlinear distortions are specifically introduced by the mobile phone and that all other elements in the acquisition chain are purely linear processing.

Characterisation is based upon the comparison of measured loudspeaker responses to a specially crafted test input. As in [21–23], and as illustrated in Fig. 3, measurements were performed using an exponential sine-sweep input signal ($sin[w_{var}]$) covering a frequency range between $f_1 = 20$Hz and $f_2 = 4$kHz. The test signal is 10s in duration and is sampled at a frequency of 8kHz.

The loudspeaker response can be seen as a composite set of components containing not only the traditional linear impulse response $\mathbf{g}_1$ but also the separate responses for each order of harmonic distortion $\mathbf{g}_p$ for $p > 1$. The different components can be isolated through the method describe in [23]. The measured set of impulse responses $\mathbf{g}_p = [g_p(0), g_p(1), \cdots, g_p(L-1)]^T$ for $p \geq 1$ are not directly the Volterra Kernels $\mathbf{h}_p$ in Eq. 3, but these can be computed easily according to the procedure described in Section 2.3.

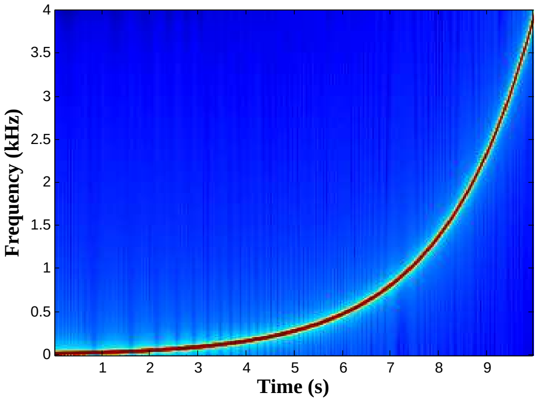

Fig. 3: Spectrogram of the exponential sine-sweep test signal, $sin[w_{var}]$.

2.2. Equalisation

The recordings described above were collected in a non-anechoic acoustic booth. While reverberation is low, recordings reflect both loudspeaker behaviour and room acoustic effects. The measured harmonic responses $\mathbf{g}_p$, $p \geq 1$ are thus equalized in order to suppress the influence of the room impulse response (RIR):

$$\mathbf{g}_{eq,p} = \mathbf{h}_{eq} * \mathbf{g}_p;\ p \geqslant 1 \tag{4}$$

where $\mathbf{h}_{eq} = [h_{eq}(0), h_{eq}(1), \cdots, h_{eq}(L_{eq}-1)]^T$ is an RIR equalisation filter. It is estimated according to the approach described in [25].

2.3. Parametrisation

After computing $\mathbf{g}_{eq,p}$, the loudspeaker response to an exponential sine-sweep ($sin[w_{var}]$) input signal can be represented as:

$$\mathbf{x}_{out} = \sum_{p=1}^{N} \mathbf{g}_{eq,p} * sin[p\omega_{var}] \tag{5}$$

The response of the polynomial Hammerstein model to the same exponential sine-sweep is given by:

$$\mathbf{x}_{out} = \sum_{p=1}^{N} \mathbf{h}_p * sin^p[\omega_{var}] \tag{6}$$

The relation between Eqs. 5 and 6 is discussed in [21] which also describes a procedure to compute the Volterra Kernels $\mathbf{h}_p$ from the measured responses $\mathbf{g}_{eq,p}$.

2.4. Application

Fig. 4 shows the practical model topology. Input signals undergo two-fold filtering by: (i) a non-linear loudspeaker response (LSR) and (ii) a room impulse response (RIR). The latter allows the application of the non-linear model in any acoustic environment different to that used for practical measurements.

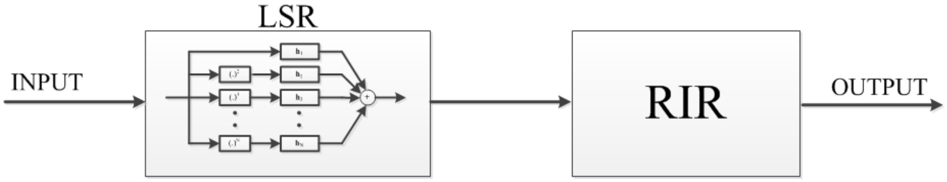

Fig. 4: Application of the non-linear loudspeaker model. Input signals are processed according to the non-linear loudspeaker response (LSR) and a room impulse response (RIR). The LSR is the Hammerstein model in Fig. 1

.

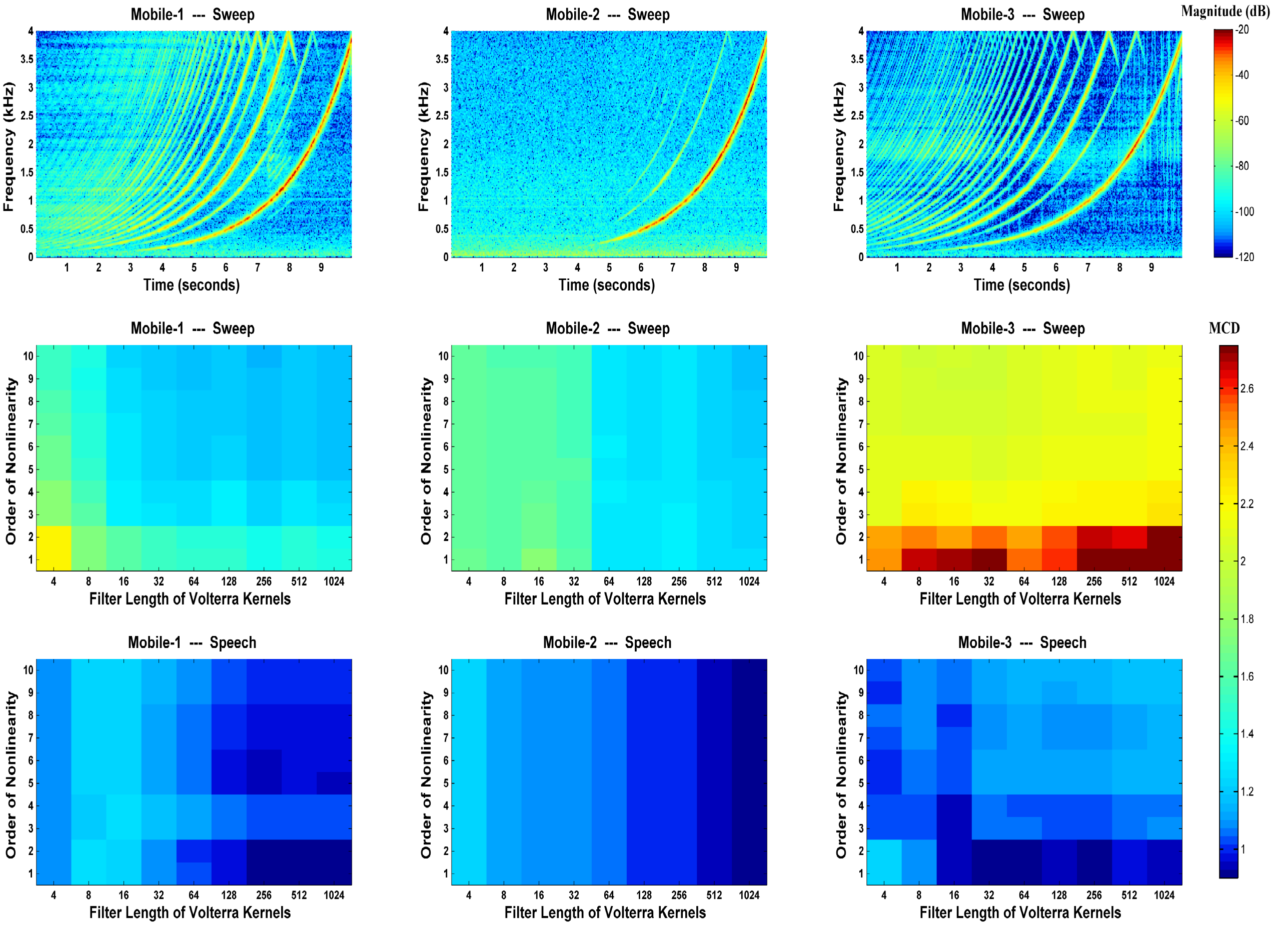

Fig. 5: An illustration of non-linear characterisation and model performance. The first row illustrates the response of each of three devices to the exponential sine-sweep input signal. Rows two and three illustrate the performance of the resulting non-linear model to sine-sweep and real-speech input signals respectively. Results shown for different orders of non-linearity N (vertical axes) and Volterra kernel lengths L (horizontal axes).

3. EVALUATION

This section illustrates non-linear behaviour in the case of exponential sine-sweep input signals which are used to characterise a particular mobile device. The performance of derived non-linear models are then assessed first, with sinusoidal inputs and second, with real-speech signals. The evaluation involves three different mobile devices with different, miniature loudspeakers used in hands-free mode at maximum volume. In applying the non-linear model, for all experiments reported below, we used the same vocal booth RIR measured in model estimation. It has a fixed length of 1024 taps at an 8kHz sampling frequency.

3.1. Device characterisation

The response of all three devices to the exponential sine-sweep input signal is shown in the form of spectrograms in the top row of Fig. 5. The 1st and particularly the 3rd device (left and right columns in Fig. 5) exhibit significant non-linear distortion; spectrograms show additional higher order harmonics in addition to the input exponential sine-sweep input signal. The non-linearity is furthermore asymmetric; odd-order harmonics are more significant than even-order non-linearities. We note that some independent studies [17,26] have reported similar observations. In contrast, the second device exhibits comparatively less non-linear distortion.

3.2. Assessment

Polynomial Hammerstein models of all three devices were measured using the procedure described in Section 2. Model performance was assessed by comparing model and real loudspeaker outputs for a common input signal. Two different input signals were used: (i) the same exponential sine-sweep signal used in the experimental procedure and (ii) a real-speech signal. Real loudspeaker signals were recorded at the ear of the mannequin using the same experimental test-bed shown in Section 2. All signals are pulse code modulation signals sampled at 8 kHz.

Performance is assessed objectively in terms of the Mean Cepstral Distance (MCD) between the real recorded signals and model estimates:

$$\begin{aligned} CD(m) &= \sqrt{\sum_{N_s} \left[C_{x_{real}}(m) - C_{x_{model}}(m)\right]^2} \\ MCD &= mean(CD) \end{aligned} \tag{7}$$

where $C_{x_{real}}(m)$ and $C_{x_{model}}(m)$ are the column vectors of cepstral coefficients from the real recorded signal x_{real} and the model output x_{model} of the m^{th} frame respectively. N_s is the length of the frame. Lower MCDs indicate that the model more accurately reflects the real measured outputs.

3.3. Results

Results for each of the three devices are illustrated in Fig. 5. The middle row shows results for the exponential sine-sweep input signal whereas the lower row shows results for the real-speech input signal. In all cases, results are shown for different orders of non-linearity N (vertical axes) and different Volterra kernel lengths L (horizontal axes). Blue colours illustrate lower MCDs whereas red colours indicate higher MCDs.

For satisfactory performance, the order of non-linearity N should be high enough to capture the principal sources of non-linear distortion, i.e. the most dominant harmonics. The Volterra kernel filter length L should be sufficiently high so as to capture accurately both linear and non-linear loudspeaker behaviour. Both parameters are however a compromise between performance and computational efficiency.

3.3.1. Exponential sine-sweep input

The response of each device to the exponential sine-sweep input signal is illustrated in middle row of Fig. 5. For the 1st and 3rd devices, the MCD is higher for lower values of N, irrespective of the number of filter taps L. The MCD nonetheless decreases with increasing N. This behaviour is not observed for the 2nd device where, in any case, the level of non-linear distortion is comparatively low. It is nonetheless reassuring that there is negligible change in model accuracy for increasing (overestimated) N. For the 1st and 2nd devices, the MCD decreases as the kernel length L increases. However, for the 3rd device, with a value of $N > 2$ performance is relatively stable for varying L. One possible explanation for such behaviour is that the highest order of significant non-linearity exceeds that of the model ($N = 10$). Since the 3rd device exhibits non-linearity greater than 10th order, N is not sufficient in this case to reduce the MCD. Accordingly, values of $N > 10$ would be needed where processing capacity allows.

3.3.2. Real-speech input

Results for real-speech inputs are illustrated in the last row of Fig. 5. Due to aliasing caused by the static non-linearity modelling, MCD values are generally lower for speech than sine-sweep inputs. For the 1st device, the best performance is obtained for lower values of N and higher values of L. For the 2nd device, performance is best for higher values of L but is independent of N. For the 3rd device performance is best in the case of $N = 1$ and values of L around 64.

These results show that, for the two cases where non-linearity is significant, the linear model ($N = 1$) outperforms the non-linear model ($N > 1$) in the case of real-speech inputs. One explanation for this behaviour lies in the wider variation in amplitude for speech signals compared to sine-sweep signals; lower amplitude speech signals may provoke significantly less non-linear distortion. It is also possible that the model obtained from the system response to sine-sweep signals is overly simplistic. Whereas the sine-sweep signal consists in a single sinusoidal frequency at any instant, speech has a far more complex spectral density whereas the model neglects inter-spectral influences.

4. CONCLUSIONS AND PERSPECTIVES

This paper investigates the use of polynomial Hammerstein models for the characterisation and modelling of non-linear loudspeakers. The Volterra kernels, which characterize the non-linear system, are empirically measured and then used to predict the response of three difference devices. Whereas validation with the same sine-sweep input signals used for characterisation shows the potential, the model yields worse performance than a conventional linear model in the case of real-speech inputs. The work highlights the challenge to model accurately the distortion introduced by non-linear loudspeakers. Further refinements are thus necessary to achieve consistent practical performance, in particular with respect to inter-spectral influences. Future work should develop new modelling strategies based on real-speech input signals rather than specially-crafted, yet artificial inputs such as those used in this work. This will allow for the full consideration of intrinsic speech characteristics and the response of non-linear systems to amplitude variations and the distribution of non-linearities across the full spectrum.

5. REFERENCES

[1] J.M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S.H. Jensen, "Linear-in-the-parameters nonlinear adaptive filters for loudspeaker modeling in acoustic echo cancellation," *J. Audio Eng. Soc*, vol. 1, 2013.

[2] M.I. Mossi, C. Yemdji, N. Evans, C. Beaugeant, and P. Degry, "Robust and low-cost cascaded non-linear acoustic echo cancellation," in *Proceedings. (ICASSP '11).*, 2011.

[3] J. Fu and W.P. Zhu, "A nonlinear acoustic echo canceller using sigmoid transform in conjunction with rls algorithm," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, Oct 2008.

[4] D.A. Bendersky, J.W. Stokes, and H. S. Malvar, "Nonlinear residual acoustic echo suppression for high levels of harmonic distortion.," in *Proceedings. (ICASSP '08).*, March 2008.

[5] L.A. Azpicueta-Ruiz, M. Zeller, J. Arenas-Garcia, and W. Kellermann, "Novel schemes for nonlinear acoustic echo cancellation based on filter combinations," in *Proceedings. (ICASSP '09).*, April 2009.

[6] J.M. Gil-Cacho, M. Signoretto, T. van Waterschoot, M. Moonen, and S.H. Jensen, "Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm," *Audio, Speech, and Language Processing, IEEE Transactions on*, Sept 2013.

[7] D. Comminiello, M. Scarpiniti, L.A. Azpicueta-Ruiz, J. Arenas-Garcia, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, July 2013.

[8] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proceedings. (ICASSP '05).*, March 2005.

[9] M.I. Mossi, N.W.D. Evans, and C. Beaugeant, "An assessment of linear adaptive filter performance with nonlinear distortions," in *Proceedings. (ICASSP '10).*, March 2010.

[10] M.I. Mossi, C. Yemdji, N.W.D. Evans, C. Hergoltz, C. Beaugeant, and P. Degry, "New models for characterizing mobile terminal loudspeaker distortions," in *IWAENC*, 2010.

[11] M. Soria-Rodriguez, M. Gabbouj, N. Zacharov, M.S. Hamalainen, and K. Koivuniemi, "Modeling and realtime auralization of electrodynamic loudspeaker nonlinearities," in *Proceedings. (ICASSP '04).*, May 2004.

[12] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing*, vol. 80, 2000.

[13] K. Lashkari, "A novel volterra-wiener model for equalization of loudspeaker distortions," in *Proceedings. (ICASSP '06).*, May 2006.

[14] D. Franken, K. Meerkotter, and J. Wassmuth, "Passive parametric modeling of dynamic loudspeakers," *Speech and Audio Processing, IEEE Transactions on*, Nov 2001.

[15] H.-K. Jang and K.-J. Kim, "Identification of loudspeaker nonlinearities using the narmax modeling technique," *J. Audio Eng. Soc*, 1994.

[16] J.F. Barrett, *Lectures on Nonlinear Systems*, Technical University Eindhoven, 1976.

[17] W. Klippel, "Loudspeaker nonlinearities - causes, parameters, symptoms," in *Audio Engineering Society Convention 119*, Oct 8-10 2005.

[18] F.X.Y. Gao and W.M. Snelgrove, "Adaptive linearization of a loudspeaker," in *Proceedings. (ICASSP '91).*, Apr 1991.

[19] L.K. Gudupudi, C. Beaugeant, N.W.D Evans, M.I. Mossi, and L. Lepauloux, "A comparison of different loudspeaker models to empirically estimated non-linerities," in *HSCMA '14, May, 2014*, Nancy, FRANCE.

[20] Marc Rebillat, Romain Hennequin, Etienne Corteel, and Brian F. G. Katz, "Prediction of harmonic distortion generated by electro-dynamic loudspeakers using cascade of hammerstein models," in *Audio Engineering Society Convention 128*, May 2010.

[21] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential sweptsine signal," *Instrumentation and Measurement, IEEE Transactions on*, Aug 2010.

[22] A. Farina, A. Bellini, and E. Armelloni, "Non-linear convolution: A new approach for the auralization of distorting systems," in *Audio Engineering Society Convention 110*, May 2001.

[23] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Feb 2000.

[24] ITU, "Objective measuring apparatus, head and torso simulator for telephonometry," in *ITU-T P.58: Terminals and Subjective and Objective Assessment Methods*, Aug 1996.

[25] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *The Journal of the Acoustical Society of America*, 1979.

[26] J.M. Gil-Cacho, T. Van Waterschoot, M. Moonen, and S.H. Jensen, "Study and characterization of the odd and even nonlinearities in electrodynamic loudspeakers by periodic random-phase multisines," in *Audio Engineering Society Convention 127*, Oct 2009.