MODELING AND REAL-TIME AURALIZATION OF ELECTRODYNAMIC LOUDSPEAKER NON-LINEARITIES

Marcelo Soria-Rodríguez
Moncef Gabbouj

Tampere University of Technology
Institute of Signal Processing
Tampere, Finland
msoriaro@vodafone.es

Nick Zacharov, Matti S. Hämäläinen
Kalle Koivuniemi

Nokia Research Center
Speech and Audio Systems Laboratory
Tampere, Finland
nick.zacharov@nokia.com

ABSTRACT

The non-linear modeling of an electrodynamic speaker is studied and its use for real-time auralization of an arbitrary sound source is considered. First, the dominant observed non-linear behavior is presented. Then, a black-box approach to the system identification task is used seeking a generalized procedure applicable to any loudspeaker. In this respect, a separate identification scheme of the linear and nonlinear characteristics of the loudspeaker is proposed. Some physical insight is used later in the model, which causes the loss of some of the desired generality. The model quality is finally assessed through subjective listening tests, and the results are presented.

1. INTRODUCTION

Modeling a loudspeaker is a challenging task that has attracted the attention of researchers [1, 2]. Of special interest is to model the non-linear behavior that appears when the loudspeaker is driven at high amplitudes. Most of the approaches have tried to obtain a model of the dominant non-linearities, in order to compute an inverse filter to pre-compensate the incoming signal to the loudspeaker. Only some non-linear effects have been considered in those cases. Here we are facing a somewhat different problem. Our aim is to obtain a model which will behave perceptually as the physical loudspeaker under any possible working conditions, with any arbitrary input signal. This model must allow the real-time simulation of the loudspeaker.

Little research has been done in this direction, with the most notable work being by Klippel [3]. In that work, a physical modeling approach was used. This approach was first introduced in [1] and has been widely used [4]. It allows for precise simulation of a number of properties of the real system, but it has the disadvantage of complexity, need for special input signals to identify and estimate the model's parameters, and the burden that it would mean to rewrite the differential equations that govern the loudspeaker if the need or desire to model another extra characteristic arises.

In this paper we use a general black-box non-linear system identification approach, which tries to reflect the complexity of the physical system with a parsimonious model. More precisely, a polynomial NARMAX (Non-linear Autoregressive Moving Average with eXogenous input) model is employed. It uses input-output measurement data to identify the real system. The data was collected in several recording sessions at Nokia Research Centre in Tampere, Finland.

The rest of this paper presents the main non-linear types of behavior encountered during the experimental sessions, the proposed approach to modeling these effects, and the results of subjective listening tests which compared the measurements of the physical driver with the model's output.

2. MODELING CHALLENGES

Next, we review the sources of non-linearities in a loudspeaker [1].

2.1. Non-linearities in the motor part

- When the loudspeaker is driven with a constant current, the force on the voice coil depends on its position. This is due to the force factor $\int Bdl$ being a function of the voice-coil excursion.
- The self-inductance of the voice-coil depends on its position, yielding a reluctance force proportional to the squared current $F_x = \frac{1}{2} i^2 \frac{dL(x)}{dx}$ (x: voice coil excursion).
- The voltage across the self-inductance is proportional both to the time derivative of the current and to the derivative of the self-inductance's own value: $U = L(x)\frac{di}{dt} + i\frac{dL(x)}{dx}\frac{dx}{dt}$.

2.2. Non-linearities in the mechanical part

- The force versus displacement curves of the loudspeaker spider and outer rim show some hysteresis.
- The voice-coil has a limited excursion range, beyond which clipping and "bottoming" effects may occur. The drive level necessary for reaching this limited excursion shall be referred to as V_{exLim}.
- Subharmonics may be generated at the loudspeaker cone when the drive level is very high.

2.3. Non-linearities in the sound radiation

- Adiabatic distortion may occur since the volume compression is not proportional to the pressure, but rather follows the relation pV^{γ} =constant.
- Doppler distortion in the high frequencies appears as a consequence of a low-frequency excursion. This kind of distortion can be neglected in practice.

2.4. Dominant non-linearities

The dominant non-linearities observed in the experiments were harmonics generation (both integer and non-integer, the latter occuring only at very high drive levels), ”bottoming” distortions and frequency-dependent gain compression .

3. PROPOSED APPROACH

A polynomial NARMAX approach [5] has been used to model the non-linear behavior of the loudspeaker, along with three linear filters L1, L2 and L3 (see figure 1 for the model structure). A brief description of the polynomial NARMAX model is now given, followed by an explanation of the proposed loudspeaker model topology.

3.1. Polynomial NARMAX

The general NARMAX model has the following expression:

$$\begin{aligned} y(t) \;&=\; f(y(t-1),\dots,y(t-n_y),u(t-1),\dots,u(t-n_u), \\ &\quad e(t-1),\dots,e(t-n_e)) + e(t) \end{aligned} \tag{1}$$

where $y(t), u(t), e(t)$ are the system's output, input and noise, respectively, and $f(\cdot)$ is a non-linear mapping from the regressors output to the system's output space. We can choose this non-linear mapping to be a polynomial, and parametrize it in the following way:

$$y_i(t) = \sum_{j=1}^{n_i} \theta_{ij} x_{ij} + e_i(t), \quad i = 1,\dots,m \tag{2}$$

This expresses a polynomial of degree L_i, with θ the parameter vector that needs to be estimated, and where

$$\begin{aligned} n_i \;&=\; \sum_{j=0}^{L_i} n_{ij}, \;\; n_{i0} = 1 \\ n_{ij} \;&=\; \frac{n_{ij-1}\left[\sum_{k=1}^{m}(n^i_{y_k} + n^i_{e_k})\sum_{k=1}^{r} n^i_{u_k} + j - 1\right]}{j} \\ &\quad j = 1,\dots,L_i \end{aligned} \tag{3}$$

and the $x_{ij}(t)$ are monomials of degrees 0 to L, each consisting, initially, of delayed outputs, inputs and noises (a degree 0 in the monomials corresponds to a constant term).

In our case, we cannot include terms which include $e(t)$ in the model, as these represent the modeling error, which is unknown when *simulating* the loudspeaker with an arbitrary input. Also, including terms which contain past outputs of the system will lead to stability problems, as the system is identified using the *measured* output, but in the real-time scenario we do not have access to that information. Therefore, we use a characterization of the NARMAX structure with $n_y = n_e = 0$ in (1).

3.2. Loudspeaker identification

Identification is done using input/output measurements with white gaussian noise input. In order to identify the non-linear effects, we propose a separate identification scheme:

1. Obtain a linear model using a low-level input. This will be the linear part of the final model (L1).
2. Obtain a second linear model using a high-level input. Use this model to simulate a different part of high-level noise, and obtain the simulation residuals with the measured output. Those residuals are supposed to contain all the non-linear effects that have appeared with the high-level noise.
3. Using the residuals as the measured output and the corresponding measured input, identify the non-linear model.

For the identification of the non-linear model's parameter vector (θ in (2)), we used an orthogonal forward-regression technique introduced in [5]. In this method, auxiliary orthogonal regressors are constructed to estimate the parameters from them. This ensures unbiased estimates, and allows model structure selection made at the same time. To do so, the algorithm computes how much does each term reduce the unexplained variance of the output variable. Selecting at each step the terms that maximize that reduction, model structure selection is performed. This selection is of capital importance: a full model set of high order polynomials using few input lags can have tens of thousands of terms, which conflicts with our real-time processing target.

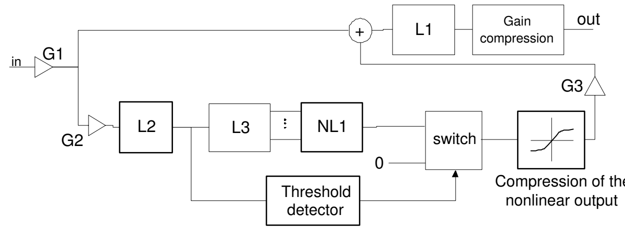

Fig. 1. *Proposed model topology*

3.3. Model topology

Figure 1 shows the proposed model topology. We present a parallel structure, where the input signal is filtered through a linear path and a non-linear path. In the following, we explain the function of each block.

3.3.1. *Linear filter L1*

This filter models the loudspeaker's impulse response. Both the input signal and the output of the non-linear path are filtered with L1, which gives the characteristic timbre of each loudspeaker.

3.3.2. *Linear filter L2*

This filter models the dependency of V_{exLim} (see §2) with frequency. Therefore, its output will be higher at the frequencies closer to the resonance frequency.

3.3.3. *Non-linear filter NL1*

The polynomial filter NL1 models some of the non-linear effects, such as harmonics, and also participates to model one of the distortions described in §2.2. In the latter case, the filter NL1 will yield a high peak when its input reaches V_{exLim}.

3.3.4. Threshold detector

When the input signal reaches V_{exLim}, this block sets the output of NL1 to zero from the next sample until two zero crossings are detected in the input (i.e., until the voice coil comes back to the same point after doing a full cycle). This creates an impulse-like peak in the signal that will be filtered by L1 (the peak created by NL1 followed by the zeros forced by this block), thus inserting at that point its impulse response. This is used to model ”bottoming” distortions (see §2.2).

3.3.5. Gain compressor and gain blocks G1-G3

At the end of the processing chain, a compressor is placed to model the change in the gain of the loudspeaker as the input level is increased. The global gain compression is measured using increasing amplitude noise, and this compression curve is implemented in this block. In the physical loudspeaker, this compression varies with frequency, but it is implemented as an equal gain reduction for all frequencies in the model.

The blocks G1-G3 are used to calibrate the levels of the linear part and the non-linear part.

3.3.6. Antialiasing filter L3

This filtering is performed to avoid the generation of frequencies above half the sampling frequency f_s by the non-linear polynomial. Here, two approaches are possible: one lowpass filter with its cutoff frequency set to $(0.5f_s)/l$, where l is the maximum degree of the non-linear polynomial; or a filter bank, with one filter per each degree present in the polynomial, such that each of them filters out the frequencies above $(0.5f_s)/i$, where $i = 1 \ldots l$ is the degree present in the polynomial. The output of the ith filter is fed to the terms of degree i in the polynomial.

3.3.7. Compression of the output of the non-linear block

If the input reaches a very high level, the polynomial can output an excessively high value. This saturation function takes care of limiting the output, such that it will compress only when an excursion limit is reached and the output value of the block NL1 is beyond some reasonable value.

3.3.8. Short discussion

The use of blocks such as L2 and the threshold detector was motivated by the impossibility of simulating such behavior with the polynomial NARMAX scheme. In that approach, since only input/output measurements are used to identify the model, a behavior such as the distortion described in §2.2 is impossible to predict. Using the mentioned blocks causes the proposed modeling approach to lose part of its generality, as it will be valid only for a given family of loudspeakers with similar behavior.

The real-time processing goal requires the use of short filters. FIR implementation was chosen for L1, as the non-linear phase distortion introduced by IIR schemes was found to be audible when processing the peaks derived from excursion-limited distortions. IIR filters can be used for L3 and L2. These filters require less coefficients than the FIR type for the same magnitude of the frequency response, which helps meeting the real-time requirement. This goal has been met with a PIII@500 MHz computer (the model processed an input audio file in less time than its length).

Factors	**Degrees of freedom**	**Levels**
1. MODEL	4	• Linea • Linpo • Full1 • Full2 • Measured
2. LOUDSPEAKER	5	• 6 typical mobile phone transducers in typical acoustics
3. LEVEL	1	• Low (peak = V_{exLim}) • High (peak = 2 × V_{exLim})
4. PROGRAM	2	• Speech • Music (Classical) • Music (Pop)
5. LISTENER	19	20 listeners

Table 1. Summary of listening test experimental factors and levels

4. PERCEPTUAL EVALUATION

4.1. Design

In order to evaluate the perceptual performance of the non-linear modeling techniques proposed a listening test was designed.

The primary aim of the study was to develop a means of adequate perceptual modeling and auralising the linear and non-linear characteristics of a loudspeaker and associated acoustics. In this test, the reference point which all models were compared to was chosen to be the measured output of the loudspeaker when driven with the original stimuli. The ITU-T P.800 degradation category rating scale (DCR) [6] was considered suitable in this case.

A test was designed with five factors: MODEL, LOUDSPEAKER, LEVEL, PROGRAM, LISTENER, the details of which are presented in table 1. The LEVEL factor was defined to excite the loudspeaker at a low level, effectively only exciting the linear performance of the device, and at a high level for excitation of non-linearities. The MODEL factor spanned four different topologies: the one described in §3.3 (**full1**), and three models which implement a set of the blocks shown in figure 1 (**full2** lacks L3; **linpo** lacks L2, L3 and the compression functions; **linea** is only the block L1, that is, a linear model). The LOUDSPEAKER factor comprised 6 different configurations of loudspeakers in several different acoustics that are typically encountered in mobile phone devices. For each program item, the DCR test was performed comparing all stimuli against the reference. The test was implemented in the GuineaPig test system [7] providing presentation of stimuli and gathering subject response data. The user interface is presented in figure 2. Listening tests were performed over the period of one week in the Listen2 test facility [8]. The listening space provides low noise (NR15) and controlled reverberation acoustic (< 280 ms above 500 Hz), conditions which are ideal for critical listening tests of this nature. The audio for the test was stored using 16 bit 48 kHz WAVE files and reproduced from a Silicon Graphics workstation via an ADAT optical digital output, a Studer D19 24 bit DAC, and an active calibrated source (CalSo) loudspeaker. All stimuli were aligned for equal RMS reproduction level and reproduced at 65 dBA at the listening position.

20 subjects were employed for the test, all of whom formed part of the Nokia Research Center listening panel. As such all subjects were known to have normal hearing and a degree of auditory acuity and reliability in rating as evaluated through the Generalised Listener Selection (GLS) procedure [9].

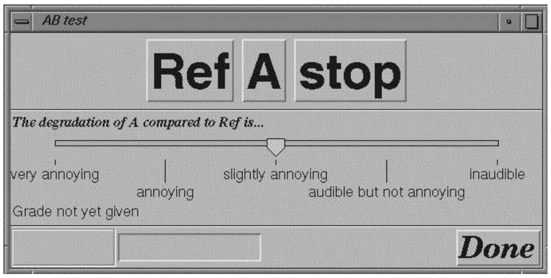

Fig. 2. GuineaPig [7] DCR [6] listening test user interface.

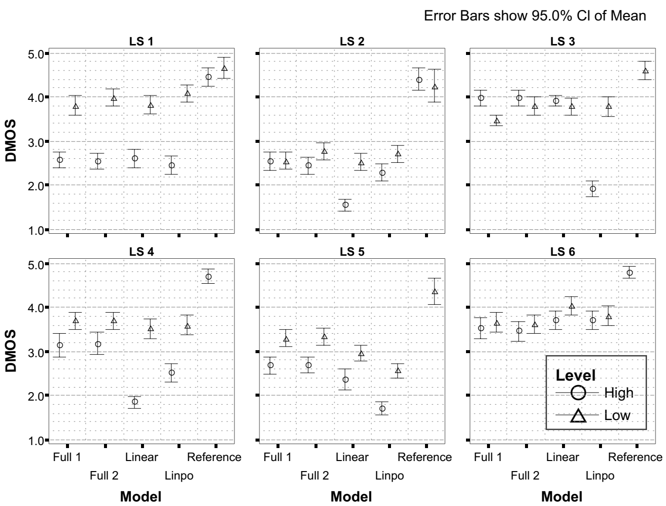

Fig. 3. Degradation mean opinion score (DMOS) values with 95% confidence intervals for all subjects averaged across PROGRAM

4.2. Analysis of the results

The data was collected and considered for analysis employing an analysis of variance (ANOVA) method. The data was tested and found to meet the basic ANOVA assumptions both in terms of normality of distribution of the raw data and also residuals (post ANOVA). A 3-way ANOVA was performed using type III sum of squares, resulting in a model explaining 71 % of the data variance. All main, 2- and 3-way interaction were found to be significant $> 95\%$ level. The five most contributing factors were: LEVEL ($F = 888.062, p < 0.000$), LOUDSPEAKER ($F = 259.892, p < 0.000$), MODEL ($F = 219.320, p < 0.000$) and 2-way interactions: LOUDSPEAKER*LEVEL ($F = 66.065, p < 0.000$) and LEVEL*MODEL ($F = 36.976, p < 0.000$).

The results of the analysis are well summarised in figure 3. In all cases the modeled performance is inferior to the measured reference stimuli. For LOUDSPEAKER configuration LS3 and LS6, there is little difference between the high and low input drive levels, illustrating that for configurations where there are limited non-linearities, the performance of all models is quite consistent. In cases LS1,2,4,5 significant non-linearities appear in the stimuli. Here we start to see varying performance between models. In several cases the **linea** or **linpo** models perform quite poorly, as clearly the modeling of non-linearities is not well achieved. It can be generalised, by averaging across the DMOS data for PROGRAM and LISTENER, that for low level stimuli (dominantly linear) all MODELS perform in a similar manner with DMOS grade ~ 3.5. When non-linearities occur at higher input levels, the **linea** and **linpo** models are found to be inferior (DMOS ~ 2.5 & 2.9 respectively). Models **full1** and **full2** perform best overall in these cases with DMOS values of ~ 3.0.

5. CONCLUSIONS

A non-linear modeling scheme has been proposed, where a polynomial NARMAX type of model is used to partially describe the physical system's dynamics. Some non-linear effects cannot be modeled within that framework, and extra elements have been proposed which model certain physical characteristics of the loudspeaker. The final model structure is thus a gray-box type approach, losing part of the desired generality. The perceptual performance of these models is moderately good, providing DMOS scores of 2.6–4.0 depending on the complexity of the stimuli. Ideally it would be desirable for the performance to exceed DMOS value 4.0 in all cases, thus some further refinement to the model is required to achieve consistent high performance.

6. ACKNOWLEDGMENTS

The authors are endebted to Andrew Bright of Nokia Research Center for his technical input and assistance to this work.

7. REFERENCES

[1] A.J.M. Kaizer, "Modeling the nonlinear response of an electrodynamic loudspeaker by a volterra series expansion," *J. Audio Eng. Soc.*, vol. 35, no. 6, June 1987.

[2] W. Klippel, "Nonlinear large-signal behavior of electrodynamic loudspeakers at low frequencies," *J. Audio Eng. Soc.*, vol. 40, no. 6, June 1992.

[3] W. Klippel, "Speaker auralization: Subjective evaluation of nonlinear distortion," http://www.klippel.de/pubs/nonlinaura.pdf.

[4] W. Frank, R. Reger, and U. Appel, "Realtime loudspeaker linearization," in *IEEE Winter Workshop on Nonlinear Digital Signal Processing*, 1993.

[5] S.A. Billings, S. Chen, and M.J.Korenberg, "Identification of mimo non-linear systems using a forward-regression orthogonal estimator," *Int. J. Control*, vol. 49, no. 6, 1989.

[6] ITU-T, *Recommendation P.800, Methods for subjective determination of transmission quality*, International Telecommunications Union, Telecommunications Standardization Sector, 1996.

[7] J Hynninen and N Zacharov, "GuineaPig - A generic subjective test system for multichannel audio," in *Proceedings of the Audio Engineering Society 106th Int. Conv.* Audio Eng. Soc., 1999.

[8] M Kylliäinen, H Helimki, N Zacharov, and J Cozens, "Compact high performance listening spaces," in *Proceedings of Euronoise*. Euronoise, 2003.

[9] V-V Mattila and N Zacharov, "Generalized listener selection (GLS) procedure," in *Proceedings of the 110th Convention of the Audio Engineering Society*, Amsterdam, Holland, 2001.