Estimation of Diagonal Volterra Kernels of an Audio System During Normal Operation with Multiple Least Mean Squares Adaptive Filters

Daniel Pinardi
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
daniel.pinardi@unipr.it

Angelo Farina
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
angelo.farina@unipr.it

Marco Binelli
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
marco.binelli@unipr.it

Andrea Toscani
Dept. of Engineering and Architecture
University of Parma
Parma, Italy
andrea.toscani@unipr.it

***Abstract*—The usage of Complete Volterra Kernels for emulating the nonlinear behavior of sound systems has been investigated for decades. Due to the computational load, the real-time implementation is typically limited to second order distortion and not feasible for higher orders. This is usually unsatisfactory for audio systems in which the disturbing distortions occur mostly at orders three and five. The same authors of this work already solved the problem with the Diagonal Volterra Kernels technique, which allowed to model arbitrarily high distortion orders. The estimation of the coefficients was obtained by exciting the system with an Exponential Sine Sweep signal. However, the result was often suboptimal since the signal reproduced by the sound system is usually different from a sinusoid. In this paper, a new method for estimating the Diagonal Volterra Kernels coefficients is proposed, by employing any music, noise or speech signal being played by a sound system in real-time. Multiple Least Mean Square algorithms are used to estimate the coefficients up to the 5th distortion order, thus allowing to emulate the nonlinearities of a typical audio system.**

Keywords—Adaptive filter, Diagonal Volterra Kernels, LMS algorithm, nonlinear audio system, real-time processing, Simulink

I. INTRODUCTION

The assessment of the nonlinear behavior of loudspeakers and other components of a sound system has always been of primary importance [1]–[7], both as a criterion for evaluating the sound degradation and for faithfully emulating the device under test (DUT), which can be also a music instrument [8] or even a room [9]. Among the various approaches, the one based on a Complete Volterra Kernel (CVK) model [10] proved good results. However, it entails a very hard computational load: being L the number of taps required for modeling the memory of the system and N the considered distortion order, the result is a hyper-cubic multidimensional matrix having N dimensions, each of them of length equal to L. The computational load quickly becomes too high for performing real-time convolution of an audio signal with distortion orders higher than two.

The problem has been solved with the introduction of Diagonal Volterra Kernels (DVK) [11]. This allows for modelling arbitrarily high distortion orders, represented by unidimensional vectors with L coefficients, instead of the previously described hyper-dimensional matrices. The technique of exciting the system with an Exponential Sine Sweep (ESS) [12] gained great success for the measurement of the linear response as well as the high distortion orders of the sound systems, which can be used to feed a DVK model. Such technique was subsequently improved with the introduction of a synchronized sine sweep [13] and a phase-correction applied in post-processing [14].

The DVK approach with ESS excitation is a realistic and efficient model of the harmonic distortion content when the DUT is stimulated by signals having marked tonal components, as the pure tone traditionally employed for measuring the Total Harmonic Distortion (THD). However, real sound systems are usually employed with signals very different from a pure tone, such as music but also broadband noise, e.g., in the case of Active Noise Control (ANC) systems installed on modern cars [15]. Under these conditions, multiple nonlinearities are stimulated and the DVK model obtained with the ESS technique is unable to adequately describe them, since it was estimated for sinusoids only.

In this work, a new, more reliable method for estimating the DVK coefficients is presented. It makes use of multiple Least Mean Square (LMS) algorithms [16] and the real signals played by the sound system, rather than a sinusoidal test signal. The method can be employed for characterizing the nonlinear behavior of several sound devices during normal operation such as loudspeakers, guitar distortions, amplifiers, or electrodynamic shakers. The effectiveness of the presented model is proved in two experimental measurements performed on a laptop loudspeaker and a resonance speaker. The paper is organized as follows: Section II describes the proposed method, Section III presents the experimental verification, and conclusions are summarized in Section IV.

II. REAL-TIME MULTIPLE LMS MODEL

The reference model for this work was proposed in [17]. The approach employs an LMS algorithm fed with the input signal x to estimate the output signal y of the system. The residual error e is obtained as:

$$e(n) = d(n) - w^T(n)x(n) \tag{1}$$

where $d(n)$ is the input signal $x(n)$ filtered by the acoustic transfer path, $w(n)$ is the adaptive filter, ^{T} denotes transpose and n is the time index. When the algorithm converges, the following relation ideally holds:

979-8-3503-1104-4/23/

$$y(n) = w^T(n)x(n) \tag{2}$$

thus, resulting in $e(n) = 0$. The adaptive filter $w(n)$ is a Finite Impulse Response (FIR) filter of L samples length, updated at each sample n, as follows:

$$w(n+1) = w(n) - \mu x(n)e(n) \tag{3}$$

where μ is the convergence coefficient (or step-size). However, in real systems only the estimation of the linear response y' of the system is obtained, and the residual error e, containing nonlinearities and noise, is always $e(n) \neq 0$. Instead of employing just one LMS for estimating the linear response, the coefficients of the DVK model are estimated with a number N of LMS, one for each kernel. In this way, the residual nonlinear signal e is assessed with a series of LMS algorithms in cascade, each fed with the input signal raised to the power N of the corresponding distortion order.

A Single-Input Single-Output (SISO) implementation of the model was developed in Simulink environment (Fig. 1). However, such configuration can be extended to any number of input and output channels, hence a Multiple-Input Multiple-Output (MIMO) configuration is also possible. It was opted to use the Normalized LMS algorithm [16], which adjusts the convergence coefficient μ by weighting it in proportion to the reference signal power, as:

$$\mu = \frac{\beta}{x'^T(n)x'(n)} \tag{4}$$

where n is the time index, x is the input reference signal, ^{T} denotes the transpose, and β is a constant between 0 and 2.

The system being analyzed is supposed to produce nonlinear harmonic distortions mainly at orders three and five, with negligible distortion at even orders. Hence, the target is to estimate three Diagonal Volterra Kernels as Impulse Responses (IR), at order one (linear IR), three, and five. A first LMS, fed with x, estimates the linear response of the system and provides a residual error e. A second LMS, fed with x^3, provides the output signal y'' that estimates the residual error e of the linear LMS. By subtracting the output of this block, y'', from the residual error signal e, a second residual signal e' is obtained, which contains the information not described by the linear and the cubic models. The third LMS, fed with x^5, aims to further reduce the residual signal e', providing as output y'''. Hence, a new residual error signal can be obtained by subtracting the output signal of the last LMS from the previous residual error signal, and additional LMS blocks could be added to estimate them. Therefore, the approach can be extended to any desired order. In the present work, it was opted to limit the number of LMS to three, since the 3rd and 5th nonlinear distortion orders are usually the most relevant in real audio systems.

One can note the model is provided with a time-domain scope for each LMS showing the real-time waveform of each kernel. An additional time-domain scope and an output block allow see and store on disk in real-time the output system signal y, the linear estimation signal y', the nonlinear distortion estimation signals y'' (3rd order) and y''' (5th order), and the residual errors signals e, e', e''.

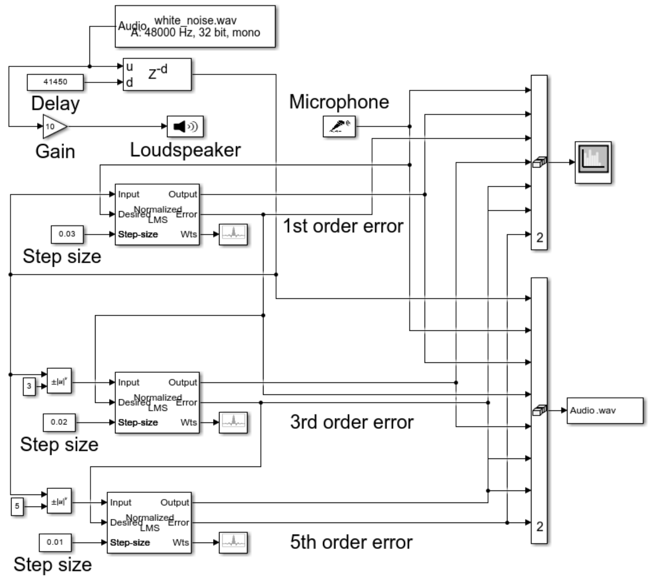

Fig. 1: Simulink model for DVK coefficients estimation with linear response and distortion orders 3 and 5.

III. Experimental Verification

This section describes the experimental measurements performed to assess the effectiveness of the proposed model, by using a laptop audio system and a resonance shaker.

A. Measurement setup

The Matlab/Simulink model was configured to use the integrated sound card of the laptop being employed to run it, interfaced with Simulink through the Windows DirectSound driver, operating at *48 kHz, 16 bits* with a buffer size *S=1024* samples. The three LMS blocks have been set for storing filters of length $L = 512\ samples$. Each of them is provided with a constant block to define the step-size of the algorithm: $\mu = 0.03$ for the linear one, $\mu = 0.02$ for the 3rd distortion order, and $\mu = 0.01$ for the 5th distortion order.

The model is provided with a tunable delay block (Fig. 2). It is required to compensate the latency of the system, which is the sum of two components:

- Internal delay, given by the sound card and the software.
- Acoustic path delay, given by the time-of-flight of the sound wave from the source (e.g., a loudspeaker) to the receiver (e.g., a microphone).

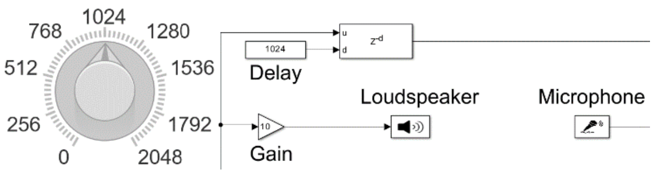

Fig. 2: Delay compensation block.

The delay is expressed in samples and must be adjusted so that the time domain response of the system under test falls within the LMS filters of L samples length. In Fig. 3, the linear estimation performed with a LMS algorithm of the integrated sound system of a laptop is shown in time domain, after having applied a correct delay compensation.

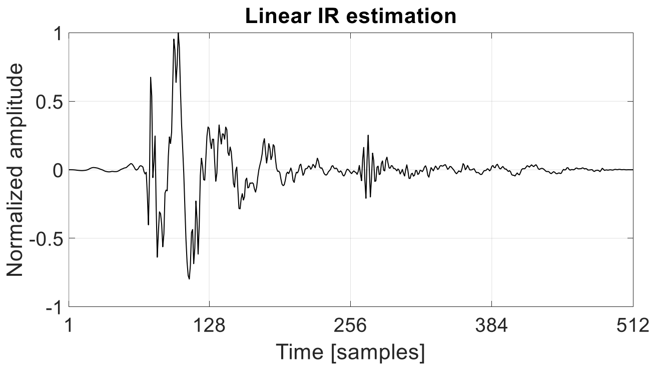

Fig. 3: Linear estimation performed by a LMS algorithm with correct delay compensation.

Regarding the test signals, each experiment was repeated twice. First, the ESS excitation technique was used, hence a *10 s* sine sweep from *20 Hz* to *20 kHz*, with *0.1 s* of fade-in and fade-out, was directly played through the system. This can be done with any Digital Audio Workstation (DAW) software that allows to synchronously play and record the test signal through the sound system. The recorded signal is then convolved with the associated inverse filter, namely the inverse sweep, to get the time domain IR. The ESS technique allows for identifying the linear response and the harmonic distortion components of the system in time domain, which allow for computing the DVK using the method first described in [11], providing at the same time a significant increase of the Signal to Noise Ratio (SNR), up to 70 dB. In Fig. 4, one can observe the linear response and the 3rd and 5th order distortion components obtained with the traditional ESS technique.

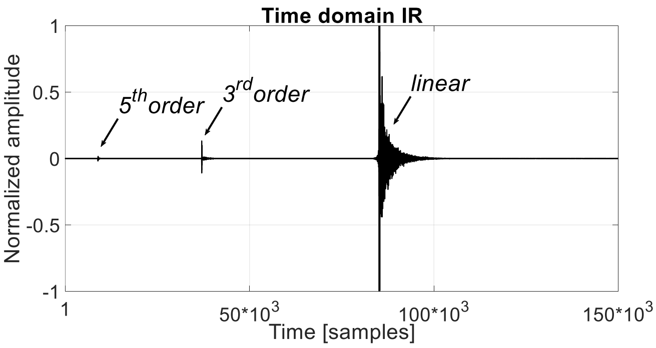

Fig. 4: Time domain IR of a sound system obtained with ESS.

The excitation signal employed to feed the real-time Simulink model was a white noise of *120 s* length to ensure the perfect convergence of the system, which in any case was accomplished after a few seconds, as it can be observed in Fig. 5.

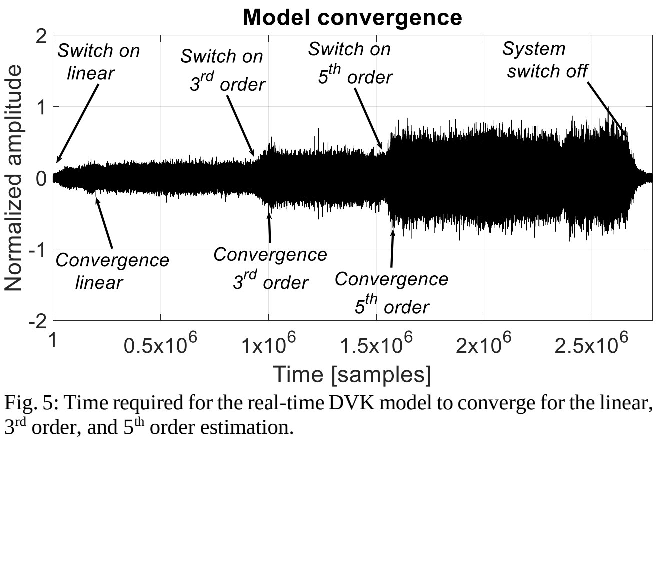

Fig. 5: Time required for the real-time DVK model to converge for the linear, 3rd order, and 5th order estimation.

B. Results for the Integrated Sound System of a Laptop

In the first experiment, one loudspeaker (left) and one microphone (left) of the integrated sound system of a laptop were used to play and record the test signal. First, the ESS technique was employed, and then the Simulink model was used to estimate the DVK coefficients. The real-time DVK model was processed twice, by enabling only the linear part, and then by also including the 3rd and 5th distortion orders. Results can be seen in Fig. 6 and Fig. 7, respectively, in terms of Sound Pressure Level (SPL) spectra. The spectra were calculated by averaging multiple Fast Fourier Transform (FFT) blocks, having an FFT size of 2^{14} samples each, overlapped by 75% with a Hann windowing. The following spectra are shown:

- Output signal recorded by the microphone (solid line).
- Estimated signal with LMS (dotted line). Only linear part and including 3rd and 5th distortion orders in Fig. 6 and Fig. 7, respectively.

The residual error is significant when the linear part only is estimated, particularly towards higher frequencies. Instead, the estimation improves in the entire frequency range after activating also the second and the third LMS blocks to estimate the residual error of the linear part.

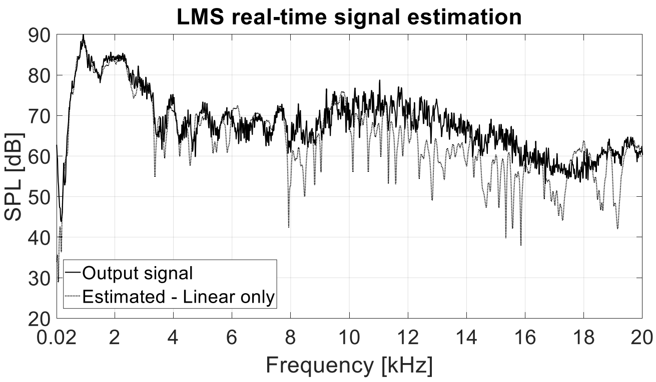

Fig. 6: Spectra of output signal and estimated signal with linear LMS only when playing a white noise through internal loudspeaker. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

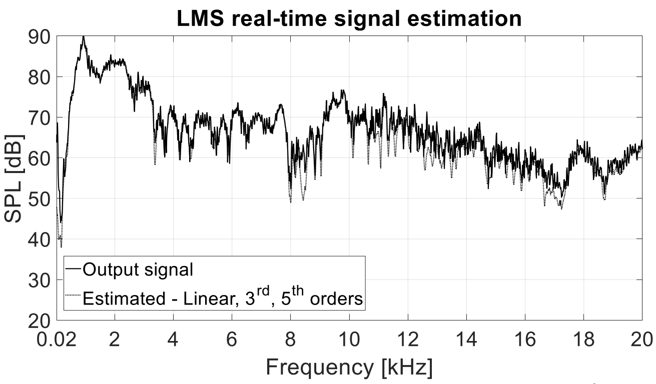

Fig. 7: Spectra of output signal and estimated signal with linear, 3rd order, and 5th order LMS when playing a white noise through internal loudspeaker. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

Eventually, the ESS and the real-time LMS modeling techniques have been compared too, in terms of linear response estimation (SPL spectra in Fig. 8), 3rd distortion order estimation (SPL spectra in Fig. 9), and 5th distortion order estimation (SPL spectra in Fig. 10). Results are summarized in Table I. One can note the ESS provided similar performance with respect to the real-time LMS model when estimating the linear response, except toward high frequencies, which causes an average loss of about *1.6 dB*. Conversely, the new real-time LMS model provided significative better performances, with an improvement of about *8.8 dB* and *4.5 dB* observed at 3rd and 5th orders, respectively. As can be calculated with the Root Mean Square (RMS) values reported in Table I, the total error including 3rd and 5th distortion orders is reduced by *26.9%* with respect to the linear only estimation when the real-time LMS technique is employed, and only by *13.1%* when the ESS technique is employed. In conclusion, the LMS model allowed to reduce the residual error of the DVK estimation by *34.8%* with respect to the ESS technique.

TABLE I. AVERAGE SPL AND SOUND PRESSURE OF EACH DISTORTION ORDER ESTIMATION FOR ESS AND LMS MODELS

Distortion order	ESS [dB]	ESS [Pa RMS]	LMS [dB]	LMS [Pa RMS]
1st (linear)	72.1	0.0805	73.7	0.0968
3rd	51.7	0.0077	60.5	0.0212
5th	43.2	0.0029	47.7	0.0049

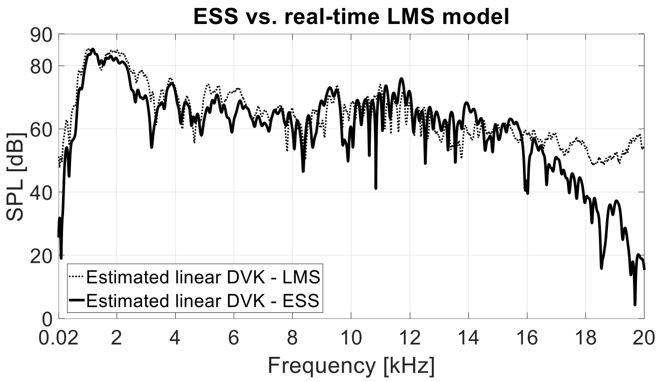

Fig. 8: Spectra of the linear DVK estimated with traditional ESS and the new real-time LMS model. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

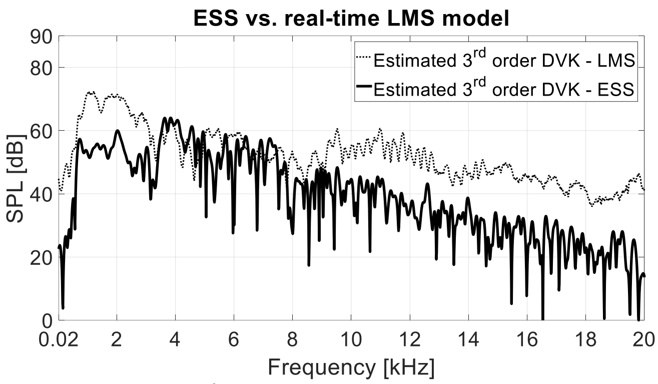

Fig. 9: Spectra of the 3rd order DVK estimated with traditional ESS and the new real-time LMS model. FFT parameters: size of 214 samples, Hann windowing, 75% overlap.

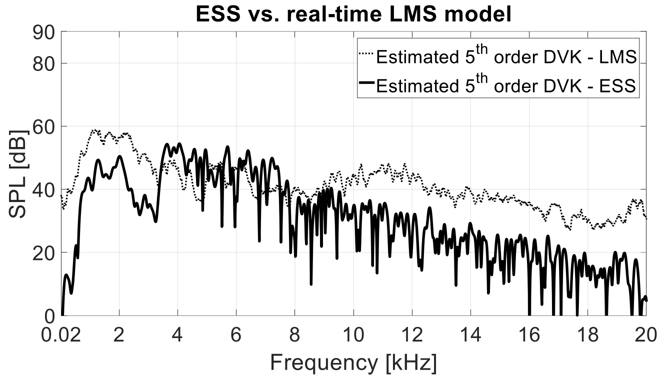

Fig. 10: Spectra of the 5th order DVK estimated with traditional ESS and the new real-time LMS model. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

C. *Result for a Resonance Speaker*

In the second experiment, the same laptop with integrated sound card and one microphone was used, but the integrated loudspeaker was replaced with an external resonance speaker, namely a shaker placed on a wooden table. Also in this case, both traditional ESS technique and the new real-time DVK model were employed. Fig. 11 shows the result for the latter approach with linear estimation only, while Fig. 12 is obtained by including the 3rd and the 5th distortion orders estimation too.

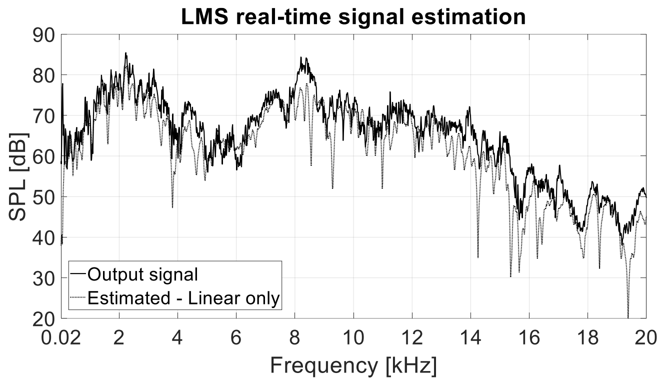

Fig. 11: Spectra of output signal and estimated signal with linear LMS only when playing a white noise through internal loudspeaker. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

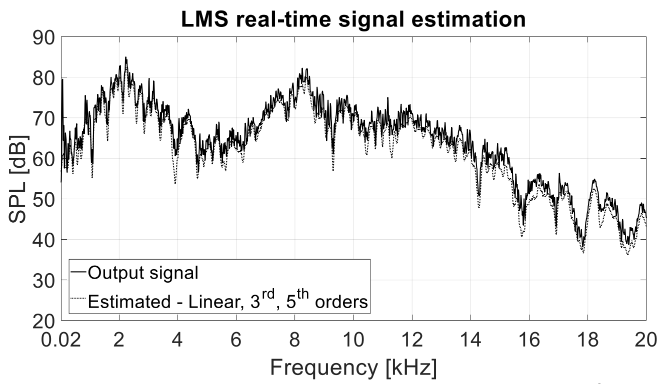

Fig. 12: Spectra of output signal and estimated signal with linear, 3rd order, and 5th order LMS when playing a white noise through internal loudspeaker. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

Eventually, the ESS and the real-time LMS model were compared, in terms of linear response estimation (SPL spectra in Fig. 13), 3rd distortion order estimation (SPL spectra in Fig. 14), and 5th distortion order estimation (SPL spectra in Fig. 15). Results are summarized in Table II. Also in this case, the ESS provided acceptable performance when modelling the linear part only (about 1.3 dB difference), with a performance reduction towards higher frequencies. The real-time LMS model provided a significative advantage in the estimation of the higher distortion orders, with an improvement of about *9.1 dB* and *10.6 dB* observed at 3rd and 5th orders, respectively. As can be calculated with the RMS values reported in Table II, the total error including 3rd and 5th distortion orders is reduced by *36.9%* with respect to the linear only estimation when the real-time LMS technique is employed, and only by *14.3%* when the ESS technique is employed. In conclusion, the LMS model allowed to reduce the residual error of the DVK estimation by *39.1%* with respect to the ESS technique.

TABLE II. AVERAGE SPL AND SOUND PRESSURE OF EACH DISTORTION ORDER ESTIMATION FOR ESS AND LMS MODELS

Distortion order	**ESS [dB]**	**ESS [Pa RMS]**	**LMS [dB]**	**LMS [Pa RMS]**
1st (linear)	77.9	0.1570	79.2	0.1824
3rd	58.3	0.0164	67.4	0.0469
5th	49.6	0.0060	60.2	0.0205

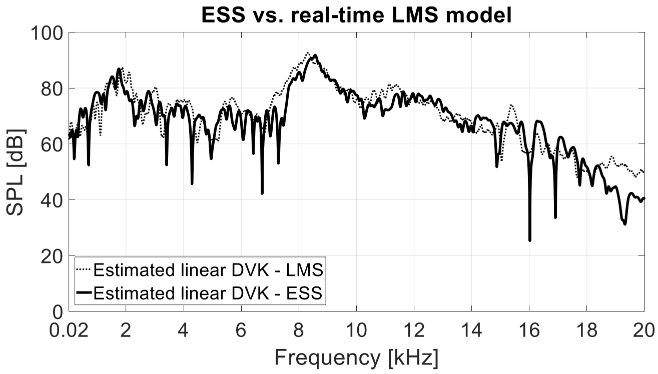

Fig. 13: Spectra of the linear DVK estimated with traditional ESS and the new real-time LMS model. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

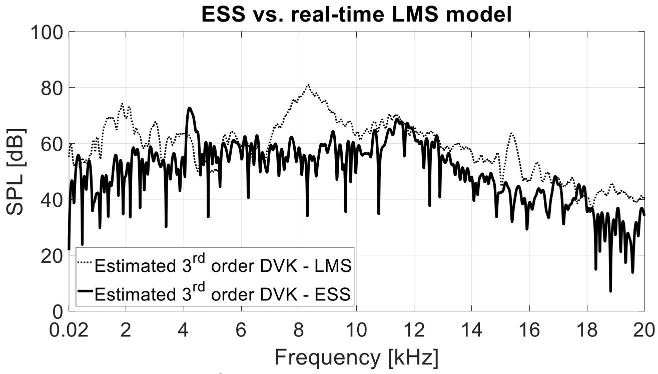

Fig. 14: Spectra of the 3rd order DVK estimated with traditional ESS and the new real-time LMS model. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

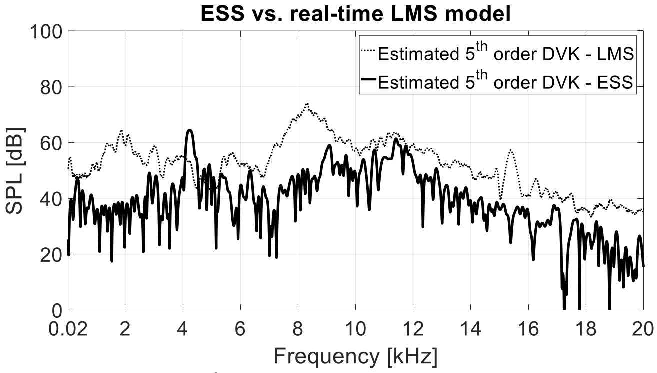

Fig. 15: Spectra of the 5th order DVK estimated with traditional ESS and the new real-time LMS model. FFT parameters: size of 2^{14} samples, Hann windowing, 75% overlap.

IV. CONCLUSIONS

A model making use of multiple Least Mean Square algorithms has been proposed for estimating in real-time the Diagonal Volterra Kernel coefficients including high distortion orders. A Simulink implementation of the model was presented, in a single-input single-output configuration for including 3rd and 5th distortion orders estimation. The presented model is scalable to a greater number of LMS blocks including higher distortion orders and to multiple-input multiple-output configurations, to operate with more than one source and more than one receiver at the same time.

Two experimental measurements were presented, one performed with the integrated sound system of a laptop computer, and one with a commercial resonance speaker. They both gave evidence that the proposed solution can significantly reduce the residual error, hence providing a better description of the nonlinearities of the system under test. The new method was compared with the traditional ESS excitation technique. It employs excitation signals more like those played by the device during real working operation (e.g., wide spectrum signals like human voice and music). In both experiments, it provided better performance with respect to the traditional one, confirming the initial hypothesis that the Volterra coefficients estimation is heavily affected by the test signal. Hence, this new technique provides a significative advancement in the emulation of the nonlinear behaviors of any audio system.

REFERENCES

[1] W. Klippel and J. Schlechter, "Distributed Mechanical Parameters Describing Vibration and Sound Radiation of Loudspeaker Drive Units," *Audio Engineering Society - 125th Audio Engineering Society Convention 2008*, vol. 1, Jan. 2009.

[2] L. Collini, A. Farina, R. Garziera, D. Pinardi, and K. Riabova, "Application of laser vibrometer for the study of loudspeaker dynamics," in *Materials Today: Proceedings*, 2017. doi: 10.1016/j.matpr.2017.06.044.

[3] M. C. Bellini, L. Collini, A. Farina, D. Pinardi, and K. Riabova, "Measurements of loudspeakers with a laser doppler vibrometer and the exponential sine sweep excitation technique," *AES: Journal of the Audio Engineering Society*, vol. 65, no. 7–8, 2017, doi: 10.17743/jaes.2017.0017.

[4] W. Klippel, "Adaptive stabilization of electro-dynamical transducers," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 1113–1117.

[5] W. Klippel, "Adaptive inverse control of weakly nonlinear systems," in *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE Comput. Soc. Press, pp. 355–358. doi: 10.1109/ICASSP.1997.599646.

[6] K. Iwai and Y. Kajikawa, "Nonlinear distortion reduction for electrodynamic loudspeaker using nonlinear filtering," in *2014 22nd*

European Signal Processing Conference (EUSIPCO), 2014, pp. 1357–1361.

[7] M. Faifer, R. Ottoboni, and S. Toscani, “A novel, cost-effective method for loudspeakers parameters measurement under non-linear conditions,” in *2008 IEEE AUTOTESTCON*, IEEE, Sep. 2008, pp. 562–568. doi: 10.1109/AUTEST.2008.4662681.

[8] L. Tronchin, “The Emulation of Nonlinear Time-Invariant Audio Systems with Memory by Means of Volterra Series,” *Journal of the Audio Engineering Society*, vol. 12, pp. 984–996, Jun. 2012.

[9] L. Fuster, M. de Diego, M. Ferrer, A. Gonzalez, and G. Pinero, “Nonlinear filtered-X second-order adaptive volterra filters for listening-room compensation,” in *2015 23rd European Signal Processing Conference (EUSIPCO)*, IEEE, Aug. 2015, pp. 669–673. doi: 10.1109/EUSIPCO.2015.7362467.

[10] M. Schetzen, “The Volterra and Wiener Theories of Nonlinear Systems,” 1980.

[11] A. Farina, A. Bellini, and E. Armelloni, “Non-Linear Convolution: A New Approach for the Auralization of Distorting Systems,” Jun. 2001.

[12] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique,” in *108th Convention*, Paris, France, 2000, pp. 1–24. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10211

[13] A. Novak, L. Simon, F. Kadlec, and P. Lotton, “Nonlinear System Identification Using Exponential Swept-Sine Signal,” *IEEE Trans Instrum Meas*, vol. 59, pp. 2220–2229, Jun. 2009, doi: 10.1109/TIM.2009.2031836.

[14] L. Tronchin and V. Coli, “Further Investigations in the Emulation of Nonlinear Systems with Volterra Series,” *J Audio Eng Soc*, vol. 63, pp. 671–683, Jun. 2015, doi: 10.17743/jaes.2015.0065.

[15] C. Belicchi *et al.*, “ANC: A Low-Cost Implementation Perspective,” in *SAE Technical Papers*, 2022. doi: 10.4271/2022-01-0967.

[16] S. M. Kuo and D. R. Morgan, “Active noise control: a tutorial review,” *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–975, Jun. 1999, doi: 10.1109/5.763310.

[17] Y. Mu, P. Ji, W. Ji, M. Wu, and J. Yang, “On the linearization of a parametric loudspeaker system by using third-order inverse Volterra filters,” in *2013 IEEE China Summit and International Conference on Signal and Information Processing, ChinaSIP 2013 - Proceedings*, Jun. 2013, pp. 570–574. doi: 10.1109/ChinaSIP.2013.6625405.