TOWARDS REPRODUCIBLE EVALUATION OF AUTOMOTIVE HANDS-FREE SYSTEMS IN DYNAMIC CONDITIONS

Marc-André Jung, Lucca Richter, and Tim Fingscheidt

Institute for Communications Technology, Technische Universität Braunschweig
Schleinitzstr. 22, D–38106 Braunschweig, Germany
{m-a.jung,l.richter,t.fingscheidt}@tu-bs.de

ABSTRACT

Reproducible evaluation of dynamic and nonlinear systems is a non-trivial problem. However, automotive speech processing algorithms such as hands-free systems have to be tested under numerous time-variant conditions in a repeatable fashion. The current way of generating time-varying echo paths, as described in ITU-T Recommendation P.1110, relies on a rotating reflecting surface in a car interior, which lacks both flexibility and reproducibility. We propose an automotive loudspeaker-enclosure-microphone (LEM) system identification approach based on the normalized least mean squares (NLMS) algorithm and a perfect sweep excitation signal. Time-variant simulations of a nonlinear system model show a significant improvement of error signal attenuation by over 7 dB, compared to a white noise excitation, also confirmed by automotive measurements. We present the necessary steps to identify dynamic automotive LEM systems to obtain traces of impulse responses for later reproducible tests of automotive hands-free systems. The method has been proposed to ITU-T standardization in focus group (FG) CARCOM.

Index Terms— NLMS system identification, automotive hands-free system, perfect sweep, simulated nonlinear system, IR database

1. INTRODUCTION

Evaluation of dynamic and nonlinear systems is a challenging task. The generation of time-variant automotive echo paths according to ITU-T Rec. P.1110 [1] by rotating a reflecting surface suffers from a lack of flexibility, reproducibility, and time efficiency. For a proper evaluation of speech enhancement algorithms, e. g. [2–6], however, a flexible and reproducible way of processing test speech data over a time-variant loudspeaker-enclosure-microphone (LEM) system model would be desired. Further development of automotive hands-free systems would highly benefit from the availability of exemplary dynamic impulse response traces from real automotive environments, provided by a database included in a new Recommendation to allow for comparability amongst different measurements and labs. This database would enable the generation of automotive test speech data, representing numerous dynamic conditions.

Whereas system identification [5, 7] is a task well understood, identification of dynamic systems is still a research topic of high demands [8–10]. For this purpose oftentimes adaptive filters are employed, whereas LMS-type algorithms convince with a lower numerical complexity compared to affine projection, recursive least squares, or Kalman algorithms [3, 11]. Excitation signals may range from noise(-like) sequences [12] over sweep signals [13], with higher energy efficiency, to perfect sequences [14–18] with an impulse-like autocorrelation function. In combining the advantages of the latter two, normalized least mean squares (NLMS) system identification with perfect sweep (PS) excitation [19, 20] shows promising results.

In our methodology we carry on Antweiler et al.'s work [20] by employing a dynamic acoustic room simulation with nonlinear processing. This accounts for the fact, that good system identification results considerably rely on the achievable signal-to-*observation*-noise ratio (SNR). The SNR strongly depends on the chosen excitation signal and its induced loudspeaker nonlinearities [13, 21, 22] at high volume. We underpin our simulation results with automotive measurements, revealing a high consistency, and present the necessary steps to acquire a dynamic impulse response trajectory for an automotive LEM impulse response database which can be used to perform reproducible, flexible, and time-efficient evaluations of automotive hands-free systems in simulated dynamic conditions.

The organization of the paper is as follows: Sec. 2 describes a discrete-time model for simulation of static and dynamic LEM system identification. In Sec. 3 simulation results are underpinned with dynamic automotive measurements, exemplarily presenting the necessary steps to acquire a dynamic impulse response trajectory. We then conclude our findings in Sec. 4.

2. IDENTIFICATION OF A SIMULATED NONLINEAR LEM SYSTEM

This section introduces a simulated nonlinear LEM system and the necessary steps for its identification. This simulated setup will serve as a ground truth to reliably evaluate the later proposed automotive dynamic LEM system identification process.

2.1. System model

Our system model (cf. Fig. 1) is based on the well-known setup of a system identification process, where the excitation signal $x(n)$ with sample index n is radiated over a loudspeaker into the acoustic enclosure to be identified—represented as linear time-variant impulse response $\mathbf{h}(n) = [h_0(n), h_1(n), \ldots, h_{N-1}(n)]^T$ with impulse response length N and transpose operator $[\cdot]^T$—thus forming a system output signal $d(n)$. Superimposed at the microphone with the observation noise signal $n(n)$ the resulting microphone signal $y(n)$ is subject to subtraction by an estimated system output signal $\hat{d}(n) = \hat{\mathbf{h}}^H(n)\,\mathbf{x}(n)$, with $\hat{\mathbf{h}}(n)$ being an estimated replica of the linear system $\mathbf{h}(n)$, $(\cdot)^H$ being the Hermitian operator, and $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-N+1)]^T$. The resulting error signal $e(n) = y(n) - \hat{d}(n)$ then is to be minimized by the adaptive

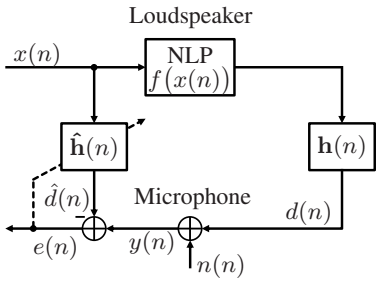

Fig. 1. Discrete-time model for identification of a time-variant system with simulated nonlinear processing (NLP) at the loudspeaker.

filter $\hat{\mathbf{h}}(n)$, in our case by making use of the NLMS algorithm, which iteratively updates the replica system's impulse response according to

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \left(\mathbf{x}(n)\, e^*(n) \right) / \|\mathbf{x}(n)\|^2, \tag{1}$$

with complex conjugate operator $(\cdot)^*$ and step size $\mu = 1$. The algorithm operates at a sampling frequency of $f_s = 16\,\text{kHz}$.

Since practical realizations oftentimes reveal considerable nonlinear behavior—mostly originating from nonlinearities of the loudspeaker—a nonlinear processing (NLP) function has been added to the loudspeaker block of the system model, thus forming a Hammerstein system [23]. This NLP is a function of the excitation signal $x(n)$ and is chosen according to [22, 24] as

$$f\big(x(n)\big) = {}^{1}\!/_{\alpha} \arctan\big(\alpha \cdot x(n)\big), \text{ with } \alpha = 0.0001, \tag{2}$$

to obtain a smooth saturation curve for higher amplitudes of the excitation signal, quantized with 16 bits.

To evaluate the quality of the identification process, an error signal attenuation measure Q according to

$$Q(n) = {}^{\mathrm{E}\{y^2(n)\}}\!/_{\mathrm{E}\{e^2(n)\}} = \frac{(1-\beta) y^2(n) + \beta \mathrm{E}\{y^2(n-1)\}}{(1-\beta)\big(d(n) - \hat{d}(n)\big)^2 + \beta \mathrm{E}\{e^2(n-1)\}} \tag{3}$$

with $\beta = 0.99$ is used. Wherever possible, the normalized system distance D according to

$$D(n) = {}^{||\mathbf{h}(n) - \hat{\mathbf{h}}(n)||^2}\!/_{||\mathbf{h}(n)||^2} \tag{4}$$

is used to measure the difference between true and replica impulse response, both of length N and $||\cdot||$ being the Euclidian norm.

2.2. Excitation signals

The type of excitation signal for a given system identification task is crucial to achieve a high SNR and—especially for time-variant scenarios—good tracking abilities to trace even highly dynamic processes. Perfect sweeps [19], as well as all perfect sequences [14–17] in general, are *perfect* in the sense, that they have an impulse-like autocorrelation function, thus leading to fast convergence [18].

Design of a PS sequence of length M in the discrete Fourier transform domain is as simple as follows:

$$\mathrm{P}(k) = \begin{cases} \exp\left(-j 4 m \pi k^2 / M^2\right), & 0 \leqslant k \leqslant \frac{M}{2} \\ \mathrm{P}^*(M-k), & \frac{M}{2} < k < M \end{cases}, \tag{5}$$

with stretch factor $m = {}^{M}\!/_{2}$, here set to equal energy distribution. Filter length and PS sequence length are chosen to $N = M = 256$ to ensure fast filter convergence whilst still providing good frequency resolution and guaranteeing periodicity with $N = M$.

Due to the sweeping character of the PS signal, its perfectness, and low crest factor, relatively high amounts of energy can be fed into the system without severe nonlinear distortions, thus leading to a high SNR [20]. Additionally, periodic repetitions are possible without transition artifacts, which further allows to increase the excitation energy. If the system to be identified is undermodeled in terms of impulse response length, the tail of the estimated impulse response—cut off after N samples—will be projected as *systematic* error at the beginning of the estimated impulse response. This systematic error appears to be more forgiving in terms of audio degradation as opposed to an *unsystematic* error, as it is observed for noise-like excitation signals (cf. Sec. 2.4 or [5, 18]).

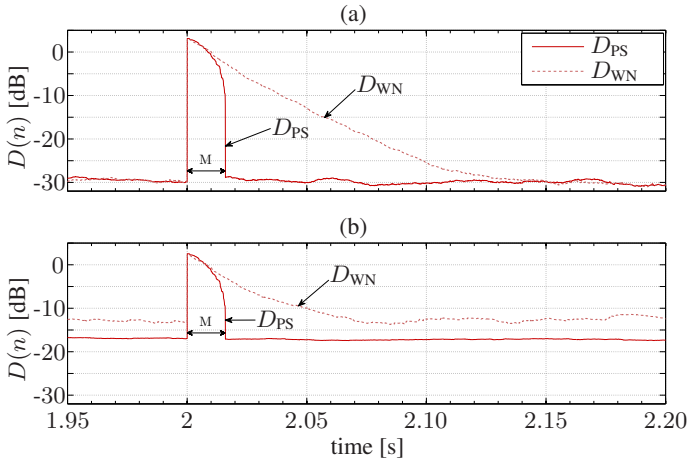

Fig. 2. Normalized system distance D for a static identification process with PS and WN excitation, $L_d = -26\,\text{dBov}$, SNR $= 30\,\text{dB}$, and $M = 256$. Switch of static impulse responses at $t = 2\,\text{s}$.
(a): without NLP, i. e. $f\big(x(n)\big) = x(n)$; **(b)**: NLP according to (2).

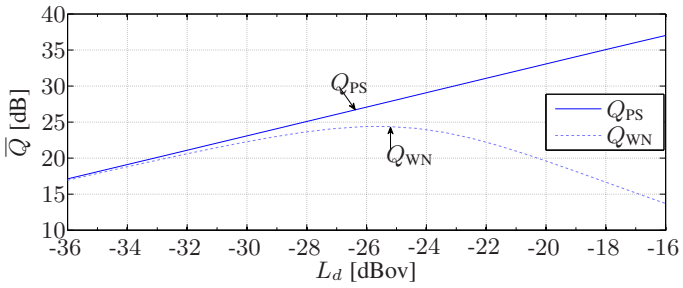

Fig. 3. Averaged simulated Q-values $\overline{Q}$ for different excitation signal types (PS and WN) and output signal levels L_d (measured at the microphone position). With NLP according to (2).

2.3. Static identification

In Fig. 2 a static system identification process according to Fig. 1 is shown, with an impulse response switch taking place at $t = 2\,\text{s}$. The underlying system—which is without NLP for the loudspeaker model in case (a) and with NLP according to (2) in case (b)—is excited with a perfect sweep sequence of length $M = 256$ and, alternatively, with white noise of equivalent energy. Subplot (a) shows that both excitation signals lead to an optimal system distance of about $D \approx -\text{SNR} \approx -30\,\text{dB}$ in the converged state. However, convergence is reached for the perfect sweep excitation signal already after one period of $M = 256\,\text{samples} \mathrel{\widehat{=}} 16\,\text{ms}$, thus representing a big advantage over the white noise excitation signal with a convergence time of about $150\,\text{ms}$. Taking NLP (2) into account, as shown

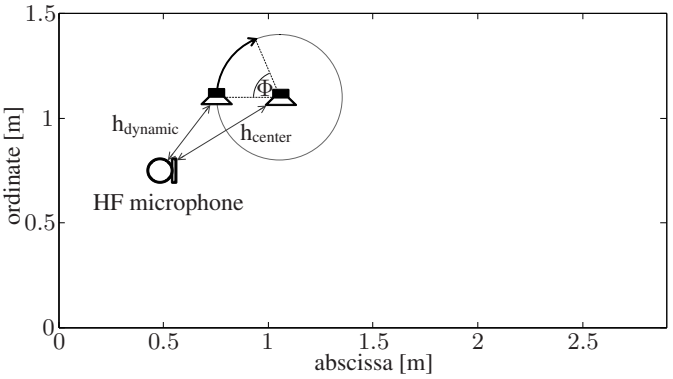

Fig. 4. Room acoustic simulation setup of a car's interior with hands-free (HF) microphone, a static center loudspeaker, and a rotating loudspeaker at a radius of 0.3 m around the static one, with Φ depicting the azimuth angle deviation from the left-most position.

in (b), the final system distance worsens to about −12 dB for WN and −17 dB for PS excitation, thus showing that PS excitation is more robust against disturbing nonlinearities with still superior convergence time compared to WN excitation.

The observed robustness against nonlinearities becomes increasingly important when higher output signal levels are desired. As shown in Fig. 3, the benefit of higher Q-values, which are expected to come with higher energetic excitation, is eminently dependent on the chosen excitation signal. To prove this, PS and WN signals have been used to excite the system shown in Fig. 1, including NLP according to (2), at output signal levels from −36 dBov to −16 dBov at the microphone. Care has been taken to ensure SNR = 30 dB for an output signal level of $L_d = -26$ dBov by adding white noise at the microphone position, remaining at a constant level of $L_n = -56$ dBov for all values of L_d. Levels have been measured according to ITU-T P.56 [25, Ch. 8].

By comparing our simulation results in Fig. 2 and Fig. 3 to real-world measurements of other labs [20, Fig. 4], a good match seems to be achieved. Therefore it can be concluded, that typical loudspeaker nonlinearities can be simulated with an NLP following (2) and that PS sequences offer a big advantage over WN as excitation signal in terms of achievable Q-values, given a specific SNR.

2.4. Dynamic identification

A simulated automotive system identification setup is shown in Fig. 4, resembling a simple cuboidal car interior with typical dimensions (2.9 m × 1.5 m × 1 m) and sound absorption properties. This setup represents a simulation analogy to the measurement setup in ITU-T P.1110 [1]. Here the time-variant echo path is realized by using a modified image method based on [26] to simulate the time-varying impulse response between a hands-free microphone at grid position 0.48 m × 0.75 m × 0.8 m and a loudspeaker, rotating at a radius of 0.3 m around the position of an imaginary co-driver (cf. [1]) at 1.05 m × 1.1 m × 0.8 m, if the ordinate is interpreted as the car's windshield plane. In order to obtain a very simple set of ground truth dynamic impulse responses, the impulse response $\mathbf{h}_{\text{dynamic}}(n)$ is combined with a static impulse response $\mathbf{h}_{\text{center}}$—between the microphone and a second loudspeaker at the aforementioned co-driver's position—to become

$$\mathbf{h}(n) = 0.1 \cdot \mathbf{h}_{\text{dynamic}}(n) + 0.9 \cdot \mathbf{h}_{\text{center}}. \tag{6}$$

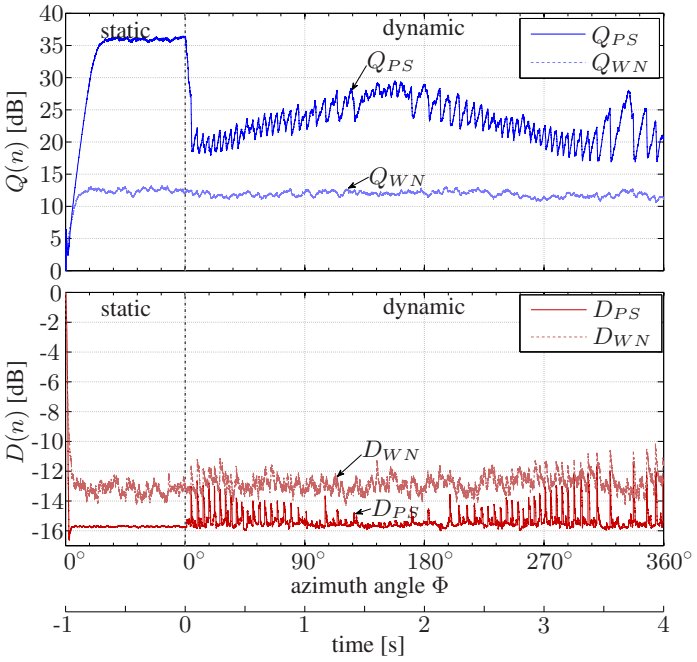

Fig. 5. Simulation results of system identification according to Fig. 4 with PS and WN excitation signals at $L_d = -16$ dBov in terms of Q-values and system distance D. SNR = 40 dB. One second of static ($\omega = 0$) identification with azimuth angle $\Phi = 0°$ and four seconds of dynamic identification with $\omega = {}^{360°}/_{4\,\text{s}}$.

This approach is based on the assumption, that most of the sound energy is conveyed over a static path. The rotation speed of the loudspeaker is $\omega = {}^{\Delta\Phi}/_{\Delta t} = {}^{360°}/_{4\,\text{s}}$, whereas the azimuth angle $\Phi = 0°$ describes the leftmost position of the rotating loudspeaker (cf. [1]).

These simulated time-varying impulse responses $\mathbf{h}(n)$ serve to feed the identification algorithm described in Sec. 2.1–2.3, with NLP according to (2) and a high output signal level of $L_d = -16$ dBov. In Addition, they constitute the ground truth for the system distance $D(n)$ calculation.

It can be seen in Fig. 5 that the results for both excitation signals nicely coincide with previous results for static identification ($\Phi = 0°, \omega = 0$). Here, PS excitation outperforms WN excitation by 24 dB in terms of Q-measure (cf. Fig. 3 for $L_d = -16$ dBov) and by about 3 dB in terms of system distance values (cf. Fig. 2 (b)). The dynamic identification process with $\omega = {}^{360°}/_{4\,\text{s}}$, shown in the time interval 0 s to 4 s, leads to somewhat worse, angle-dependent Q-measure values for the PS excitation in the range of $17\,\text{dB} < Q_{\text{PS}} < 30\,\text{dB}$, still outperforming the Q-measure values for WN excitation at $Q_{\text{WN}} \approx 12$ dB. Noteworthy are two local maxima in the Q_{PS} plot which correspond to positions of the rotating loudspeaker where it is geometrically in line with the center loudspeaker and the microphone. Taking a look at the lower subplot, it can be seen that system distance values show similar minimum values but increased maximum values compared to the static case for both excitation signals.

3. IDENTIFICATION OF A REAL AUTOMOTIVE LEM SYSTEM

Having shown the superiority of PS excitation over WN excitation in a simulated static and dynamic automotive environment, measurements in a car are conducted to investigate the portability of the

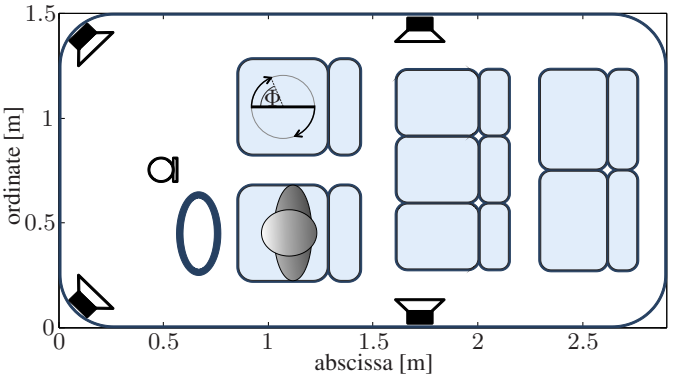

Fig. 6. Car setup with generation of a time-varying echo path according to ITU-T P.1110 [1], with hands-free microphone at rear-view mirror position and four loudspeakers. Azimuth angle $\Phi = 0^\circ$ depicts an orientation of the reflecting surface parallel to the abscissa.

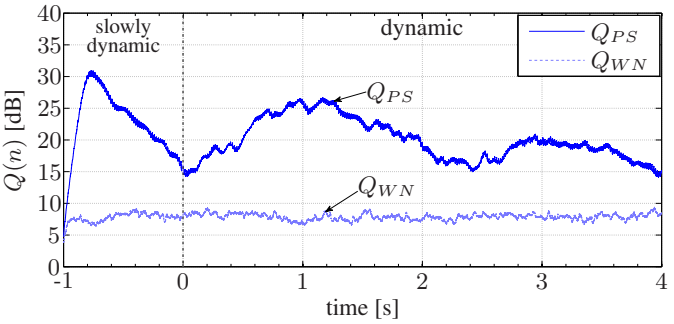

Fig. 7. Measurement results of system identification according to Fig. 6 and [1] with PS and WN excitation signals at $L_d = -16\,\mathrm{dBov}$ in terms of Q-measure values Q. After one second of slow rotation start four seconds of dynamic identification with $\omega \approx {}^{360^\circ}/_{4\,\mathrm{s}}$.

aforementioned conclusions to real-world applications.

3.1. Measurement setup

System identification measurements have been performed in a Volkswagen Touran car, with an interior setup sketched in Fig. 6 (cf. also Fig. 4). Both excitation signals, PS and WN, were played back via four internal car loudspeakers and recorded with the hands-free microphone at rear-view mirror position (grid position $0.48\,\mathrm{m} \times 0.75\,\mathrm{m} \times 0.8\,\mathrm{m}$). The normalized excitation signals were played back at high volume to achieve a good SNR. In accordance to ITU-T P.1110 [1] a piece of plywood of size $0.3\,\mathrm{m} \times 0.4\,\mathrm{m}$ was placed at the co-driver's seat and rotated with $\omega \approx {}^{360^\circ}/_{4\,\mathrm{s}}$ to generate a time-varying echo path. The initial position ($\Phi = 0^\circ$) of the board hereby again corresponds to the setup, where its surface is parallel to the abscissa. The driver's seat has been occupied.

3.2. Dynamic identification

In the described car setup an identification process of the time-variant system—created according to ITU-T P.1110, except for manual rotation of the board—was performed and the result is shown in Fig. 7. Only Q-values are provided, since now ground truth is not available to perform system distance measurements.

One second after rotation start, a speed of $\omega \approx {}^{360^\circ}/_{4\,\mathrm{s}}$ was maintained for four seconds. As it can be seen in Fig. 7, the Q-measure for WN excitation remains at a rather constant and low level of $Q_{\mathrm{WN}} \approx 8\,\mathrm{dB}$, whereas with PS excitation values for Q_{PS} from $15\,\mathrm{dB}$ to $26\,\mathrm{dB}$ can be achieved. As it already could be assumed based on the simulation results of Fig. 5, these real-world measurements also show two local maxima in the Q_{PS} plot of Fig. 7, here at about $t = 1\,\mathrm{s}$ and $t = 3\,\mathrm{s}$. These maxima positions, though somewhat misplaced due to the erratic manual rotation, belong to azimuth angles Φ of the rotating board where again specific geometrical properties are met. As a consequence, a high coherence between the Q-measure results of the simulation (Sec. 2.4) and the real-world measurements (here) can be observed.

By convolution of these acquired time-varying impulse responses $\mathbf{h}(n), n = 0, 1, 2, \ldots$, of the dynamic system with close-talk speech signals $\mathbf{x}(n)$, automotive speech signals with a high resemblance to the original room impression can be simulated. Informative subjective listening experiments showed, that impulse responses obtained by PS excitation produced far better convolution output signals compared to the excitation by WN. For WN excitation, a lot of click noises could be perceived in the convolution output signals. Furthermore, room impression differed considerably more as opposed to the PS excitation case.

NLMS system identification simulation and real-world measurements showed that low system distance and high immunity against nonlinearities can be achieved with periodic PS excitation. Therefore, this approach—NLMS identification of several time-variant automotive prototype environments to build up a database for later reproduction of these conditions via convolution in the lab—can be promoted as particularly suitable, e. g., for inclusion into a future automotive hands-free system test Recommendation. It would offer the advantage of high reproducibility and ease of use when available close-talk speech signals shall be equipped with dynamic automotive room characteristics of high realism. In so doing, a large amount of test data can be processed with different settings in the lab for dynamic automotive impulse responses of various source environments, without the effort of recording each speech file in the car individually.

Our methodology of acquiring traces of dynamic automotive impulse responses has been proposed in ITU-T focus group CARCOM to provide a database of such impulse responses along with a future ITU-T Recommendation and is currently under discussion.

4. CONCLUSIONS

Generation of time-variant speech material for automotive hands-free system testing according to ITU-T P.1110 [1] oftentimes does not lead to satisfying results, due to a lack of reproducibility, flexibility, and time efficiency. Our proposed approach detaches the dynamic room characteristics from the speech signal's content by NLMS system identification with perfect sweep sequences, thus offering a higher degree of abstraction. It represents an advancement of the state of the art by presenting a realistic automotive time-variant simulation framework with nonlinear processing, underpinning its automotive applicability with real-world measurements. By including this method into future automotive hands-free system test Recommendations, a database could be built, allowing users to rely on numerous predefined dynamic impulse response traces for flexible, reproducible, and comparable generation of automotive time-variant test data.

5. REFERENCES

[1] ITU-T, *Rec. P.1110: Wideband Hands-Free Communication in Motor Vehicles*, International Telecommunication Union, Dec. 2009.

[2] B. Widrow and P. N. Stearns, *Adaptive Signal Processing*, 1st ed. Englewood Cliffs, NJ: Prentice Hall, Mar. 1985.

[3] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, Sep. 2002.

[4] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. Chichester, Great Britain: John Wiley & Sons, 2006.

[5] M. Benesty, M. Sondhi, and Y. Huang, Eds., *Springer Handbook of Speech Processing*. Berlin / Heidelberg, Germany: Springer, 2008.

[6] E. Hänsler and G. Schmidt, Eds., *Speech and Audio Processing in Adverse Environments*. Berlin / Heidelberg, Germany: Springer, 2008.

[7] P. Eykhoff, *System Identification – Parameter and State Estimation*. New York, NY: John Wiley & Sons, 1974.

[8] L. Ljung and S. Gunnarsson, "Adaptation and Tracking in System Identification – A Survey," *Automatica*, vol. 26, no. 1, pp. 7 – 21, 1990.

[9] T. Fingscheidt and Suhadi, "Quality Assessment of Speech Enhancement Systems by Separation of Enhanced Speech, Noise, and Echo," in *Proc. of 8th Annual Conference on International Speech Communication Association (INTERSPEECH '07)*, Antwerp, Belgium, Aug. 2007, pp. 818–821.

[10] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic Measurement of Room Impulse Responses Using a Moving Microphone," *Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1636–1645, 2007.

[11] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Hoboken, NJ: John Wiley & Sons, 2004.

[12] J. Borish and J. Angell, "An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise," *Journal of the Audio Engineering Society*, vol. 31, no. 7/8, pp. 478–488, 1983.

[13] S. Müller and P. Massarani, "Transfer-Function Measurement with Sweeps," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.

[14] V. P. Ipatov, "Ternary Sequences with Ideal Periodic Autocorrelation Properties," *Radio Engineering and Electronic Physics*, vol. 24, pp. 75–79, Oct. 1979.

[15] H. D. Lüke, "Sequences and Arrays with Perfect Periodic Correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 3, pp. 287–294, May 1988.

[16] C. Antweiler and M. Antweiler, "System Identification with Perfect Sequences Based on the NLMS Algorithm," *International Journal of Electronics and Communications (AEÜ)*, vol. 3, pp. 129–134, 1995.

[17] D. Jungnickel and A. Pott, "Perfect and Almost Perfect Sequences," *Discrete Applied Mathematics*, vol. 95, pp. 331–359, 1999.

[18] C. Antweiler and G. Enzner, "Perfect Sequence LMS for Rapid Acquisition of Continuous-Azimuth Head Related Impulse Responses," in *Proc. of 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '09)*, Oct. 2009, pp. 281–284.

[19] A. Telle, C. Antweiler, and P. Vary, "Der perfekte Sweep – Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme," in *Proc. of German Annual Conference on Acoustics (DAGA '10)*. Berlin, Germany: DEGA, Mar. 2010, pp. 341–342.

[20] C. Antweiler, A. Telle, P. Vary, and G. Enzner, "Perfect-Sweep NLMS for Time-Variant Acoustic System Identification," in *Proc. of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '12)*, Kyoto, Japan, Mar. 2012, pp. 517–520.

[21] E. Hänsler and G. Schmidt, Eds., *Topics in Acoustic Echo and Noise Control*. Berlin / Heidelberg, Germany: Springer, 2006.

[22] G. Enzner, "From Acoustic Nonlinearity to Adaptive Nonlinear System Identification," in *Proc. of 10th ITG Conference on Speech Communication*, vol. 236. Braunschweig, Germany: VDE Verlag, Sep. 2012, pp. 23–26.

[23] A. Hammerstein, "Nichtlineare Integralgleichungen nebst Anwendungen," *Acta Mathematica*, vol. 54, pp. 117–176, 1930.

[24] U. Zölzer and X. Amatriain, Eds., *DAFX: Digital Audio Effects*. Chichester, Great Britain: John Wiley & Sons, Apr. 2003.

[25] ITU-T, *Rec. P.56: Objective Measurement of Active Speech Level*, International Telecommunication Union, Dec. 2011.

[26] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, 1979.