Prawda, Karolina; Schlecht, Sebastian; Välimäki, Vesa

Short-term Rule of Two: Localizing Non-Stationary Noise Events in Swept-Sine Measurements

Short-term Rule of Two: Localizing Non-Stationary Noise Events in Swept-Sine Measurements

Karolina Prawda[1], Sebastian J. Schlecht[1,2], and Vesa Välimäki[2]

[1]*Dept. of Information and Communications Engineering, Aalto University, Espoo, Finland*
[2]*Dept. of Art and Media, Aalto University, Espoo, Finland*

Correspondence should be addressed to Karolina Prawda (`karolina.prawda@aalto.fi`)

ABSTRACT

Non-stationary noise is notoriously detrimental to room impulse response (RIR) measurements using exponential sine sweeps (ESSs). This work proposes an extension to a method of detecting non-stationary events in ESS measurements that aims at precise localization of the disturbance in the captured signal. The technique uses short-term running cross-correlation as a means to estimate the instantaneous correlation between two sweep signals. Both, the detection threshold and measured correlation, are evaluated on short windows, allowing for accurate analysis of the entire signal. Additional pre-processing steps are applied to improve the robustness of the proposed technique. The approach is tested on various types of simulated and measured non-stationary noise, showing that detection errors did not exceed 23 ms. The method presented in this work increases the robustness of RIR measurements using ESS against non-stationary noise.

1 Introduction

The exponentially swept sine (ESS) [1] is currently a popular excitation signal for measuring room impulse responses (RIRs) used in various applications, from room acoustic parameters estimation [2, 3, 4, 5, 6] to head-related transfer function measurements [7]. While praised for its numerous advantages, such as high signal-to-noise ratio (SNR) [8], robustness of measurements [8, 9], and ease of harmonic distortion rejection [8, 10, 11, 12], the ESS method is vulnerable to non-stationary noise, e.g., transients. Such noise events may lead to artifacts in the deconvolved RIR and, in consequence, errors in the estimated parameters [8, 13, 14, 15, 16, 17].

Currently, two methods of detecting non-stationary noise in ESSs are established. The first one was devel-oped by Guski *et al.* and compares the energy of the noisy part of the measurement to the energy of Gaussian noise [18, 19]. The second one, introduced by Prawda *et al.*, uses the Pearson Correlation Coefficient (PCC) to measure the similarity between two signals. A low PCC value indicates the presence of non-stationary noise in one of the sweeps. This technique, called the Rule of Two (Ro2), is aimed at selecting a pair of ESS signals free from non-stationary noise when a series of measurements is performed [17].

This work extends the Ro2 by introducing short-term running cross-correlation analysis. This allows to analyze the change of the correlation value over the whole duration of the signal, in place of one PCC value for the entire sweep. The aim of the short-term analysis is to localize the non-stationary noise in the contaminated ESS.

The short-term running cross-correlation is widely used in estimation of the effects of time variance on both RIRs and sweeps [20, 21, 22]. In literature, however, the time lag of the maximal value of cross-correlation is used as an estimator of transfer-function-variation-induced changes to the signal. Here, we propose to use the maximal value itself to evaluate the similarity between the sweeps, as in [17].

The paper is organized as follows: Section 2 describes the Ro2 and detection threshold estimation, as well as introduces the short-term running cross correlation. In Section 3, the pre-processing is specified, and the validation on both, simulated and measured non-stationary disturbances is shown. Section 4 concludes the paper.

2 Methodology

This section discusses the Ro2 method and the correlation of ESS signals and proposes the method of accurate detection of non-stationary events in a sweep measurement. The pre-processing, aimed at minimizing the effect of disturbances on sweep and enhancing the detection, is discussed as well.

2.1 Rule of Two

The signal y_i captured during the ith ESS measurement can be written as

$$y_i = x_i + u_i = s * h + u_i, \tag{1}$$

where x_i is the measured sweep, s is the ESS signal, h is the RIR, the asterisk $*$ denotes convolution, and u_i is the stationary noise term.

The correlation between y_i and the consecutive measurement, y_j, is expressed as [17, 23, 24]:

$$\rho_{y_i, y_j} = \frac{\sum_n^N (x_i(n) + u_i(n))(x_j(n) + u_j(n))}{\sqrt{\sum_n^N (x_i(n) + u_i(n))^2 \sum_n^N (x_j(n) + u_j(n))^2}} \tag{2}$$

where N is the total length of the signal in samples and n is the discrete time index.

The Ro2 method selects sweep measurements free from non-stationary noise from a series of sweep recordings based on the criterion:

$$\text{if } \rho_{y_i, y_j} > \widehat{\rho}_{y_i, y_j} \text{ then } y_i \text{ and } y_j \text{ are a clean pair}, \tag{3}$$

where $\widehat{\rho}_{y_i, y_j}$ is a selection threshold, which is determined based on the influence of expected contamination—stationary background noise and transfer-function variation—during the measurement. Due to the properties of correlation being a comparative measure, Ro2 requires at least two clean signals to reliably separate sweeps containing non-stationary noise from those free from it [17]. The Ro2 method helps to automatize ESS measurements.

2.2 Selection Threshold Estimation

The selection threshold in Ro2 is determined based on the correlation decrease due to the expected contamination with stationary background noise and transfer-function variation. Here, we recall the procedures to estimate the threshold, $\widehat{\rho}_{y_i, y_j}$.

We assume that the noise terms u_i and u_j, being random, are uncorrelated with the ESS signals as well as with each other, yielding $\sum_n^N u_i u_j = 0$, $\sum_n^N x u_i = 0$, and $\sum_n^N x u_j = 0$, and transforming Eq. (2) into

$$\rho_{y_i, y_j} = \frac{\sum_n^N x_i(n) x_j(n)}{\sqrt{\sum_n^N (x_i^2(n) + u_i^2(n)) \sum_n^N (x_j^2(n) + u_j^2(n))}}, \tag{4}$$

which can be written in terms of signal energies as

$$\rho_{y_i, y_j} = \frac{E[x]}{\sqrt{(E[x] + E[u_i])(E[x] + E[u_j])}}, \tag{5}$$

assuming that the energy of each ESS signal is the same across measurements, i.e., $E[x_i] = E[x_j]$.

Additionally, if the background noise is stationary during the entire measurement series, that is, if its energy is approximately constant, $E[u_i] = E[u_j] = E[u]$, we can establish a correlation threshold based on the noise energy as in [17]:

$$\frac{E[x] + \zeta E[u]}{E[x] + E[u]} \leq \widehat{\rho}_{y_i, y_j, \zeta}, \tag{6}$$

where $\zeta = \rho_{u_i, u_j}$ symbolizes the correlation of the stationary noise terms, in case they are not perfectly uncorrelated. Here, we adopt $\zeta = -1$, considering an extreme case of noise terms being anticorrelated. Although such a scenario is virtually impossible, this presumption relaxes a very strict detection threshold posed by the assumption of $\zeta = 0$ for situations when the noise terms include harmonic content, e.g., electric humming [17].

Additionally, we need to consider that the system under test is not perfectly time-invariant, meaning that the correlation between two sweeps is affected by transfer-function variation as well. In the case of room acoustic measurements, such variations come mostly from fluctuations of atmospheric conditions—temperature and humidity—as well as air movement [10, 17, 19, 25, 26, 27, 28]. Considering that those changes affect the RIR directly, Eq. (1) becomes:

$$y_i = s * (h + v_i) + u_i, \qquad (7)$$

where v_i is the variation of RIR due to time variance. The difference between two measured signals is then

$$y_i - y_j = s * (v_i - v_j) + u_i - u_j. \qquad (8)$$

The transfer-function variation factor can be estimated from the energies according to

$$\tau = \frac{E[s * (v_i - v_j)]}{E[s * h]} = \frac{2E[s * v_i]}{E[s * h]} = \frac{2E[s * v_j]}{E[s * h]}, \qquad (9)$$

where $E[s * h]$ can be retrieved from the measurement as $E[s * h] = E[y_i] - E[s * v_i] - E[u_i]$. The selection threshold from Eq. (6) is then modified to accommodate the effect of transfer-function variation [17]

$$\widehat{\rho}_{y_i, y_j, \zeta, \tau} = \frac{\widehat{\rho}_{y_i, y_j, \zeta}}{1 + \tau/2} . \qquad (10)$$

The values of τ strongly depend on the measurement environment and the length of ESSs. For short sweeps in an controlled laboratory setting, a $\tau \ll 1$ is expected.

2.3 Short-Term Running Cross-Correlation

The equations presented in this work so far are considering the entire measured signal. Thus, we make the necessary modifications to accommodate them for short-term analysis.

Firstly, we define the short-term signal x_i at time n

$$\tilde{x}_i^n(m) = w(m - n) x_i(m) x_i(m - n), \qquad (11)$$

where w is a Hanning window, normalized so that $\sum_m w(m) = 1$, and $m = 1, ..., M$, where M is the window length. The windowing is applied to smooth the short-term cross-correlation curves.

Thus, the computation of cross-correlation from Eqs. (4) and (5) is altered to the form of windowed convolution of two signals:

$$\rho_{y_i, y_j}(n) = \frac{\sum_m \tilde{x}_{i,j}^n(m)}{\sqrt{E[\tilde{x}_i^n] E[\tilde{x}_j^n]}} . \qquad (12)$$

The detection thresholds from Eqs. (6) and (10) are modified as well, yielding

$$\frac{E[\tilde{x}^n] + \zeta E[u]}{E[\tilde{x}^n] + E[u]} \le \widehat{\rho}_{y_i, y_j, \zeta}(n), \qquad (13)$$

and in consequence, the selection threshold becomes

$$\widehat{\rho}_{y_i, y_j, \zeta, \tau}(n) = \frac{\widehat{\rho}_{y_i, y_j, \zeta}(n)}{1 + \tau/2} . \qquad (14)$$

Here, only the energy values relating to the signal itself are computed on short windows. The energy of the stationary noise as well as the transfer-function variation factor are still estimated for the entire measurement. The reason for that is the sensitivity of τ to all changes in the signal, including non-stationary noise, and the difficulty of estimating stationary noise energy when sweep is being played [18, 19].

2.4 Pre-Processing of Sweeps

The ease of use and robustness of the proposed short-term Ro2 method may be enhanced by completing simple pre-processing steps, tackling the problem of signal's SNR, and minimizing the actual effect of noise on the ESSs.

Due to the property of the sweep measurements, emitting only one frequency at a time, the presence of a non-stationary noise during the emission of the sweep does not warrant the corruption of the relevant part of the signal. For example, if a low-passed transient occurs when high frequencies are played, it will be pushed to the non-causal part of the RIR in the deconvolution process. It can be then discarded without compromising the measurement results.

The detection of non-stationary disturbances is conveniently performed on the measured sweep signal, where the artifact is compacted in time. Thus, in the present work we propose to first deconvolve the obtained sweep, truncate the resulting RIR to remove harmonic distortion and irrelevant part of the signal,
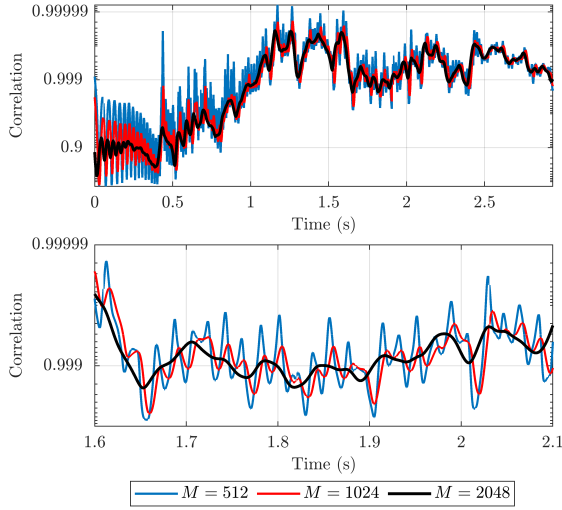
Fig. 1: Short-term running cross-correlation for a pair of clean, pre-processed sweeps, for different window lengths. (Top) the entire signal and (bottom) a snippet of 0.5 s.

and convolve them back with the input sweep for the actual detection process. That way, only the parts of the non-stationary noise that contaminate the sweep remain, increasing the robustness of localization.

As the measurement conditions are noisy, it is important to choose the right portion of the signal, i.e. with high enough SNR, for analysis. This is especially crucial since the SNR is used in establishing detection thresholds in Eqs. (13) and (14). However, setting the SNR threshold at 0 dB is not enough. After the sweep is done playing, and only the sound decay is recorded, the correlation exhibits a sudden drop whilst the SNR is still positive. Thus, the analysis of this part of the signal may prove unreliable. It is crucial, however, not to discard the decaying sound completely, as it still a part of the RIR. Therefore, in this work, we propose to use the portions of the signal where the SNR is above zero, but is a relatively low number. In the case of ESSs used in our analysis, SNR of 5 dB proved sufficient.

3 Validation

This section presents the results of non-stationary noise localization in simulated and measured scenario. The accuracy of the method is evaluated on different types of noise events.

An important parameter in the analysis is the length of the window w, as it affects the fluctuation of short-term cross-correlation curves. The said curves for a pair of measured and pre-processed ESS signals are depicted in Fig. 1, illustrating this effect. The curve for $M = 512$ is the most unstable, oscillating around more smoothed curves obtained with longer windows.

On the other hand, a large M reacts to changes in correlation slowly, and thus may conceal short non-stationary noise events or compromise the accuracy of their localization. As seen in the bottom pane of Fig. 1, some of the prominent notches are almost completely flattened for $M = 2048$. They are also shifted in time compared to the two remaining plots. Thus, in this work, the $M = 1024$ samples, which translates to 23.2 ms for the sampling rate of 44.1 kHz, is considered a reasonable compromise in terms of robustness and accuracy of the localization.

3.1 Simulations

Since there is no other established method to localize the non-stationary noise events in ESS measurements, the initial validation of the accuracy of short-term Ro2 is performed on simulations. The signals used for this part of the evaluation are sweeps measured in the variable acoustic laboratory *Arni*, located at the Acoustics Laboratory at Aalto University, Espoo, Finland. They are all 3-s-long ESS signals, and were marked as free from non-stationary disturbances by the Ro2 [17].

The clean sweeps were contaminated with different types of non-stationary disturbances: transients, elevated noise floor, and sound dropouts. RIRs measured in *Arni* and matching the decay times of the analyzed sweeps were used as transient noise. Amplified 1.5-s-long snippets of background noise from the same database were used to raise the noise floor level. Deleting between one and ten samples simulated sound dropouts. The times of appearance of the disturbances were randomized, and so were the gains for the background noise snippets and the number of dropped samples. Examples of all three types of contamination are displayed in Fig. 2.

The error measure used to evaluate the accuracy of the method was the difference between the actual time of occurrence of the noise event and the one detected by the short-term Ro2 method. The onset of the non-stationary disturbance was marked when the measured correlation descended below the detection threshold.
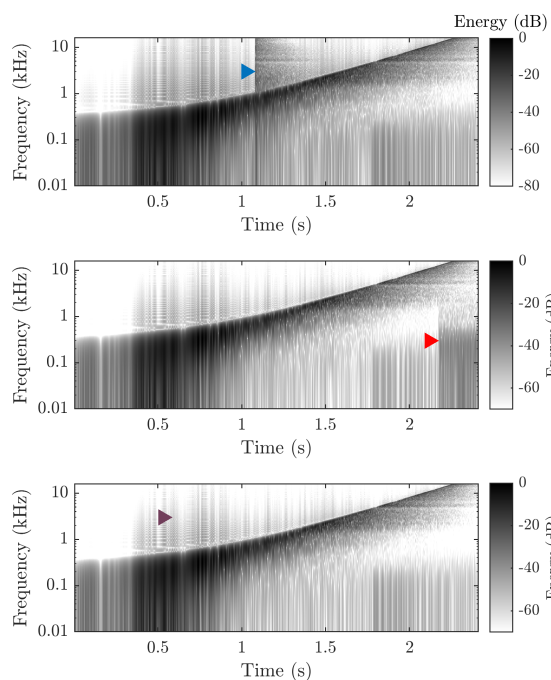
Fig. 2: Spectrograms of artificially contaminated sweeps with (top) transient, (middle) elevated noise floor, and (bottom) dropout. The arrows point to onset times of the disturbances. The signals were not subjected to pre-processing to keep the visibility of the non-stationary noise.
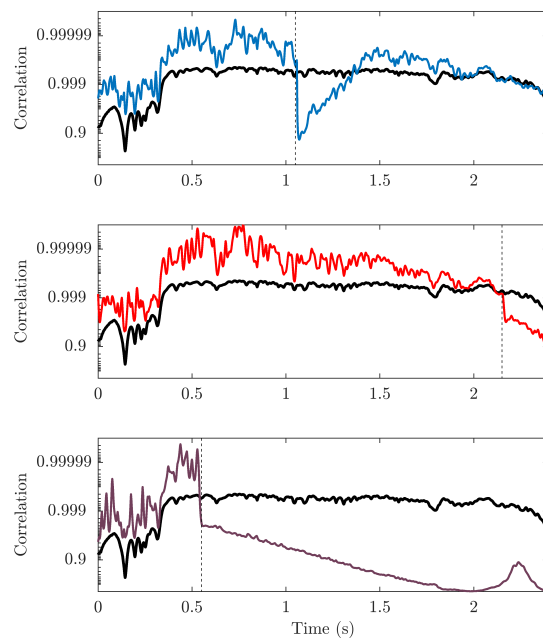

Fig. 3: Short-term cross-correlation for (top) transient, (middle) elevated noise floor, and (bottom) sound dropout. The solid black lines represent the detection threshold, whilst the colored lines depict the measured correlation for each case. The dashed vertical lines mark the onset times for detected disturbances (*cf.* arrows in Fig. 2).

The exemplary short-term running cross-correlation values for the three types of non-stationary noises are shown in Fig. 3. In each case, the presence of a disturbance marks a clear drop in measured correlation below the assumed detection threshold. In the cases of transients and added noise, the ρ_{y_i, y_j} values may return to high values when the disturbance ends (*cf.* top pane of Fig. 3). However, when the sound dropout occurs, it affects the rest of the signal, effectively dropping the correlation for the remaining duration of the sweep.

The results of the simulations are presented in Fig. 4. They show that for all the disturbances, the difference between the detected and actual onset time fits within the length of one window. The error for transients was the most consistent at around 19 ms. For the elevated noise floor, the difference varied the most, from 11 ms to about 18 ms. The error for sound dropouts was the highest, arriving close to the window length.

3.2 Measurements

The second part of the validation was performed on the ESS measurements from *Arni* database that were already flagged as containing non-stationary noise by the Ro2. Since the absence of the ground truth impedes the possibility to define an error measure, we show examples of measured non-stationary noise types and the performance of the short-term Ro2 algorithm in their presence.

The contamination by a broadband transient is shown in Fig. 5, where the top pane shows the non-stationary noise before pre-processing, and the bottom pane depicts the pre-processed ESS. Due to the process, the effect on the impulse on the sweep is decreased, and the disturbance itself is almost invisible. However, the correlation drop shown in Fig. 6 detects the remains of the transient.
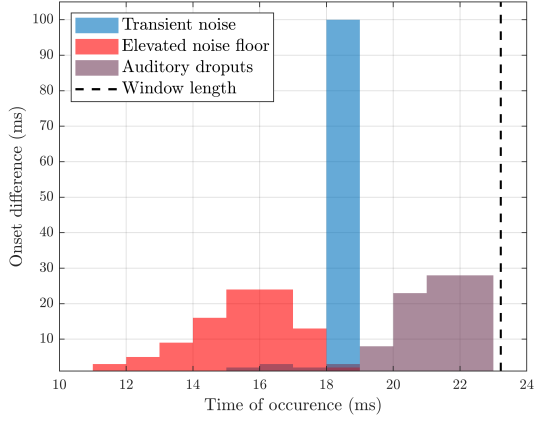
Fig. 4: The difference between the detected and actual onset times of non-stationary noise events in the simulated scenario. The dashed line marks the window length 23.2 ms used in the analysis.
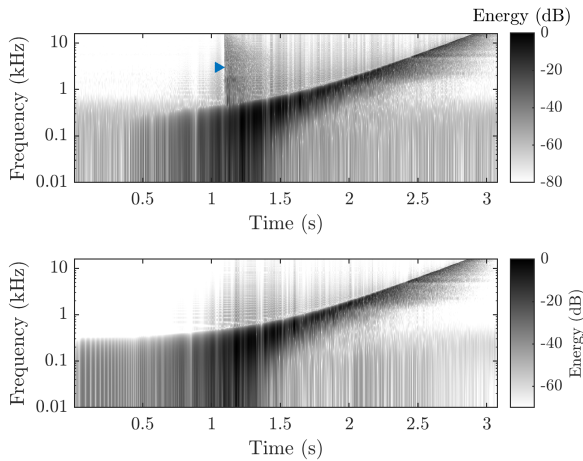

Fig. 5: Spectrograms of an ESS signal from *Arni* database contaminated with a transient noise (top) before and (bottom) after pre-processing. The arrow points to the onset of the contamination at 1.05 s.

In the case of elevated noise floor, the sweeps were contaminated by a low-frequency noise, as depicted in Fig. 7. The disturbance affects the majority of the measured signal, as shown in the bottom pane of Fig. 7, where pre-processing removed only a small part of the energy from the additional noise. Thus, the values of ρ_{y_i,y_j} stay mostly below the detection threshold for
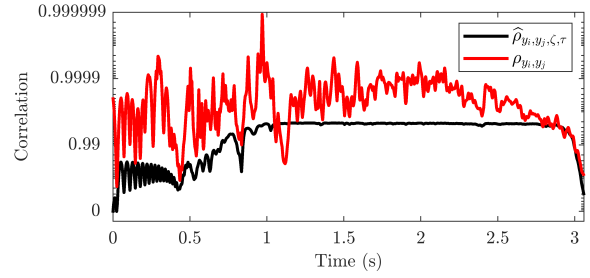

Fig. 6: Lower threshold for non-stationary noise detection (black line) and the measured short-term correlation of an ESS polluted by a transient (red). The drop in ρ_{y_i,y_j} at 1.05 s indicates a noise event, cf. Fig. 5.

almost the entire duration of the ESS, as presented in Fig. 8. Only after the contamination is gone does the correlation stay consistently above $\widehat{\rho}_{y_i,y_j,\zeta,\tau}$.

When the sound dropouts occur, the contamination is not easily visible in the spectrograms in Fig. 9. However, the correlation exhibits a sudden drop to a very low value, as shown in Fig. 10. Similarly to the simulations, the ρ_{y_i,y_j} of the rest of signal is affected, never returning above the threshold.

4 Summary and conclusions

The present work proposes an extension to an established method of detecting non-stationary noise in sweep measurements. The Rule of Two, or Ro2, was extended to accommodate precise localization of non-stationary events in the captured ESS signals, which was possible by using short-term running cross-correlation in place of a single-number correlation coefficient.

The paper presents the methodology to estimate the effect of expected contamination—stationary noise and transfer-function variation—on short-term cross-correlation values. The pre-processing including deconvolution and truncation of measured RIRs is used to minimize the effect of the non-stationary noise on the measurement. Setting a suitable SNR threshold is used to enhance the robustness of the method as well.

The short-term Ro2 was first validated on a set of simulated non-stationary noise events, with the results showing that it is capable of localizing disturbances with an accuracy of at worst the analysis window length.
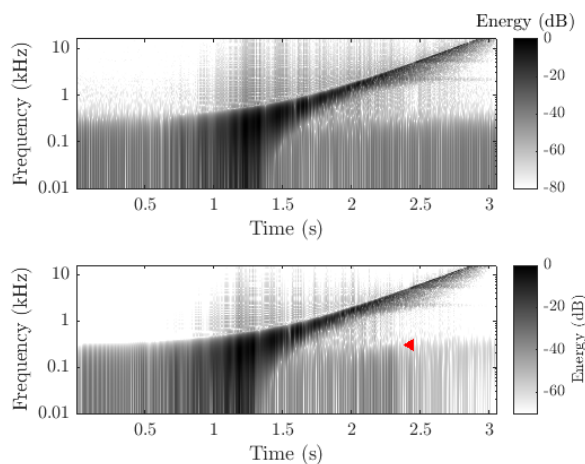
Fig. 7: An ESS signal from *Arni* database contaminated with a low-frequency noise (top) before and (bottom) after pre-processing. Since the noise contaminated the measurement from the start, its offset at 2.42 s is marked with an arrow.
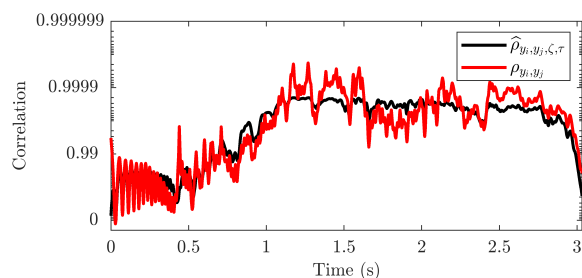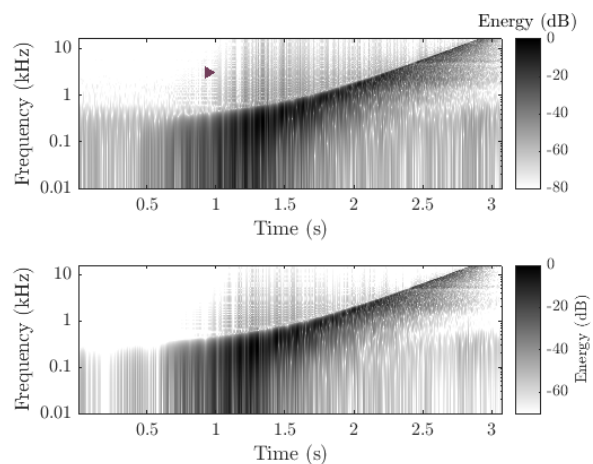

Fig. 9: An example of an ESS signal from *Arni* database containing an sound dropout (top) before and (bottom) after pre-processing. The onset of the contamination at 0.95 s is marked with an arrow.
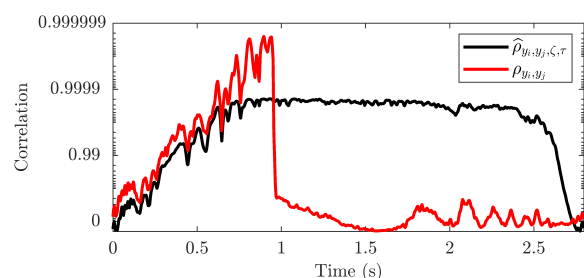

Fig. 8: Short-term correlation of an ESS containing low-frequency noise (red line) compared with the lower threshold for non-stationary noise detection (black), cf. Fig. 7.

Fig. 10: Correlation of an ESS signal corrupted by a dropout (red line) compared with the detection threshold (black), cf. Fig. 9.

The experiments on measured sweeps containing non-stationary noise are consistent with the outcome of the simulations. The results of this work prove that the short-term Ro2 is a reliable method to localize non-stationary noise in sweep measurements.

In the future, the precise localization of noise events in sweep measurements may aide in correcting noisy ESSs by non-stationary noise removal.

References

[1] Farina, A., "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th Convention*, Paris, France, 2000.

[2] Pätynen, J., Tervo, S., and Lokki, T., "Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses," *J. Acoust. Soc. Am.*, 133(2), pp. 842–857, 2013.

[3] Wenmaekers, R. H. C., Hak, C. C. J. M., and Hornikx, M. C. J., "How orchestra members influence stage acoustic parameters on five different concert hall stages and orchestra pits," *J. Acoust. Soc. Am.*, 140(6), pp. 4437–4448, 2016.

[4] Götz, G., Schlecht, S. J., Martinez Ornelas, A.,

and Pulkki, V., "Autonomous Robot Twin System for Room Acoustic Measurements," *J. Audio Eng. Soc.*, 69(4), pp. 261–272, 2021.

[5] Prawda, K., Schlecht, S. J., and Välimäki, V., "Evaluation of reverberation time models with variable acoustics," in *Proc. 17th SMC Conf.*, pp. 145–152, Torino, Italy, 2020.

[6] Prawda, K., Schlecht, S. J., and Välimäki, V., "Calibrating the Sabine and Eyring formulas," *J. Acoust. Soc. Am.*, 152(2), pp. 1158–1169, 2022.

[7] Pulkki, V., Laitinen, M.-V., and Sivonen, V., "HRTF Measurements with a Continuously Moving Loudspeaker and Swept Sines," in *Proc. AES 128th Convention*, London, UK, 2010.

[8] Stan, G.-B., Embrechts, J.-J., and Archambeau, D., "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.*, 50(4), pp. 249–262, 2002.

[9] Torras-Rosell, A. and Jacobsen, F., "A New Interpretation of Distortion Artifacts in Sweep Measurements," *J. Audio Eng. Soc.*, 59(5), pp. 283–289, 2011.

[10] Müller, S. and Massarani, P., "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, 49(6), pp. 443–471, 2001.

[11] Guidorzi, P. and Garai, M., "Impulse responses measured with MLS or swept-sine signals: A comparison between the two methods applied to noise barrier measurements," in *Proc. AES 134th Convention*, Rome, Italy, 2013.

[12] Müller-Trapet, M., "On the practical application of the impulse response measurement method with swept-sine signals in building acoustics," *J. Acoust. Soc. Am.*, 148(4), pp. 1864–1878, 2020.

[13] Farina, A., "Advancements in Impulse Response Measurements by Sine Sweeps," in *Proc. AES 122nd Convention*, Vienna, Austria, 2007.

[14] Ćirić, D., Pantić, A., and Radulović, D., "Transient noise effects in measurement of room impulse response by swept sine technique," in *Proc. 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS)*, pp. 269–272, 2011.

[15] Guidorzi, P., Barbaresi, L., D'Orazio, D., and Garai, M., "Impulse Responses Measured with MLS or Swept-Sine Signals Applied to Architectural Acoustics: An In-depth Analysis of the Two Methods and Some Case Studies of Measurements Inside Theaters," in *Proc. 6th International Building Physics Conference (IBPC)*, volume 78, pp. 1611–1616, Torino, Italy, 2015.

[16] Segerstrom, E., Lee, M.-L., and Philbert, S., "Evaluating Four Variants of Sine Sweep Techniques for Their Resilience to Noise in Room Acoustic Measurements," in *Proc. AES 147th Convention*, New York, NY, 2019.

[17] Prawda, K., Schlecht, S. J., and Välimäki, V., "Robust selection of clean swept-sine measurements in non-stationary noise," *J. Acoust. Soc. Am.*, 151(3), pp. 2117–2126, 2022.

[18] Guski, M. and Vorländer, M., "Impulsive noise detection in sweep measurements," *Acta Acust. united Ac.*, 101(4), pp. 723–730, 2015.

[19] Guski, M., *Influences of External Error Sources on Measurements of Room Acoustic Parameters*, Doctoral dissertation, RWTH Aachen University, Aachen, Germany, 2015.

[20] Satoh, F., "Acceptable temperature changes during synchronous averaging for reverberation time measuring by swept-sine method," in *Proc. 19th Int. Congr. Acoust. (ICA)*, pp. 1–6, Madrid, Spain, 2007.

[21] Postma, B. N. J. and Katz, B. F. G., "Correction method for averaging slowly time-variant room impulse response measurements," *J. Acoust. Soc. Am.*, 140(1), pp. EL38–EL43, 2016.

[22] Wang, X., *Model based signal enhancement for impulse response measurement*, Doctoral dissertation, RWTH Aachen University, Aachen, Germany, 2013.

[23] Borga, M., "Canonical correlation: A tutorial," *Online tutorial http://people.imt.liu.se/magnus/cca*, 4(5), 2001.

[24] Prawda, K., Schlecht, S. J., and Välimäki, V., "Multichannel Interleaved Velvet Noise," in *Proc. Int. Conf. Digital Audio Effects (DAFx)*, pp. 208–215, Vienna, Austria, 2022.

[25] Svensson, P. and Nielsen, J. L., "Errors in MLS Measurements Caused by Time Variance in Acoustic Systems," *J. Audio Eng. Soc.*, 47(11), pp. 907–927, 1999.

[26] Niederdränk, T., "Maximum length sequences in non-destructive material testing: application of piezoelectric transducers and effects of time variances," *Ultrasonics*, 35(3), pp. 195–203, 1997.

[27] Georgiou, F., Hornikx, M., and Kohlrausch, A., "Auralization of a car pass-by inside an urban canyon using measured impulse responses," *Appl. Acoust.*, 183, p. 108291, 2021.

[28] Vorländer, M. and Kob, M., "Practical aspects of MLS measurements in building acoustics," *Appl. Acoust.*, 52(3), pp. 239–258, 1997.