Acoust. Sci. & Tech. **43**, 3 (2022)
©2022 The Acoustical Society of Japan

ACOUSTICAL LETTER

Safeguarding test signals for acoustic measurement using arbitrary sounds: Measuring impulse response by playing music

Hideki Kawahara[1,*] and Kohei Yatabe[2,†]

[1]*Wakayama University, 930 Sakaedani, Wakayama, 640–8510 Japan*
[2]*Waseda University, 3–4–1 Ookubo, Shinjuku-ku, Tokyo, 169–8555 Japan*

(*Received 21 December 2021, Accepted for publication 25 January 2022*)

Keywords: Discrete Fourier transformation, Periodic signal, Impulse response, Spurious responses, Entertaining test sounds, Loop music

1. Introduction

We propose a simple method of measuring acoustic responses using any sound by converting them to be suitable for measurement. This method enables us to use music pieces for measuring acoustic conditions. It is advantageous to measure such conditions using test sounds that do not annoy the listeners. In addition, applying the underlying idea of the simultaneous measurement of multiple paths [1,2] provides features of practical value. For example, it is possible to measure deviations (temporally stable, random, and time-varying) and the impulse response while reproducing slightly modified contents under target conditions. The key idea of the proposed method is to add relatively small deterministic signals that sound like noise to the original sounds. We call the converted sounds as *safeguarded test signals*.

2. Safeguarded transfer function measurement

Let $x[n]$ be a periodic discrete-time signal with a period L. The convolution of $x[n]$ and the impulse response $h[n]$ of the target system yields the output $y[n]$. Because the signal is periodic, the discrete Fourier transforms (DFTs) of $x[n]$ and $y[n]$ segments (their length is L) are invariant other than the phase rotation proportional to frequency. Let $X[k]$ and $Y[k]$ represent their DFTs, where k, $(k = 0, \ldots, L-1)$, is the discrete frequency. Then, the ratio $Y[k]/X[k]$ is independent of the location of the segment. This ratio agrees with the DFT $H[k]$ of the impulse response $h[n]$, where $X[k] \neq 0$ for all k values is the condition of this relation to provide physically meaningful results.

However, this simple solution is sensitive to noise when the absolute value $|X[k]|$ is very small relative to absolute values $|H[k]|$ of other k values. We propose to limit the absolute value $|X[k]|$ to be larger than the threshold.[1] We use the following equation to derive the DFT $X_s[m]$ of the safeguarded signal $\tilde{x}_s[n]$.

$$X_s[k] = \begin{cases} \dfrac{\theta_L X[k]}{|X[k]|} & \text{for} \quad 0 < |X[k]| < \theta_L \\ X[k] & \qquad \theta_L \leq |X[k]| \end{cases} \tag{1}$$

Here, we set $X_s[k] = \theta_L$ when $X[k] = 0$. Then, we derive the safeguarded transfer function $H_s[k]$ as

$$H_s[k] = \frac{Y_s[k]}{X_s[k]}, \tag{2}$$

where $Y_s[k]$ represents the DFT of the output of the target system for the periodic test signal $\tilde{x}_s[n]$. Because the safeguarded signal $\tilde{x}_s[n]$ is periodic, we can make *the safeguarded test signal* for acoustic measurement by concatinating it as many times as required. To analyze the safeguarded transfer function, we can select the safeguarded test and the output segment anywhere, obeying one rule: the segment must have the length of exactly L samples.[2]

3. Measurement of other responses

When the target system is a linear time-invariant (LTI) system, and no observation noise exists, the calculated safeguarded transfer function $H_s[k]$ is the same irrespective of the location of the safeguarded test and the output segments. Also, $H_s[k]$ is independent of the safeguarded test signals used. However, this is not the case when measuring acoustic systems in the real world. We can use these differences of $H_s[k]$ measured at different observation locations as well as different safeguarded test signals to separate the LTI response and other spurious responses. They are signal-induced deterministic and random responses.[3]

3.1. Separation of random responses

An additive noise $d[n]$ in output observation produces a deviation term $D[k]$. We define the time-invariant response $H_{sTI}[k]$ and the squared absolute random response $|D_{sTV}[k]|^2$ by measuring the system M times.

$$H_{sTI}[k] = \frac{1}{M} \sum_{m=1}^{M} H_s^{\{n_m\}}[k] \tag{3}$$

*e-mail: kawahara@wakayama-u.ac.jp
†e-mail: k.yatabe@asagi.waseda.jp
[doi:10.1250/ast.43.209]

[1]Generally, the threshold is a function of discrete frequency $\theta_L[k]$. We use a constant value here to make explanations simple.

[2]Do not use the initial segment of length L samples (plus samples for propagation delay from the sound source to the microphone) because it does not have any preceding cycle.

[3]In addition to background and observation noises (which are independent of the source), there are source-related random noises, for example, the turbulent noise caused by strong low-frequency airflow in the bass-reflex port and the high-frequency phase modulation noise due to the Doppler effect caused by air movement [1].

$$|D_{\mathrm{sTV}}[k]|^2 = \frac{1}{M-1}\sum_{m=1}^{M}|H_{\mathrm{s}}^{\{n_m\}}[k] - H_{\mathrm{sTI}}[k]|^2 \tag{4}$$

The use of superscript $a^{\{n_m\}}$ indicates that the calculation of a uses the segment starting from the discrete-time n_m.[4] The denominator $M-1$ of Eq. (4) is for unbiased variance.

3.2. Separation of signal-dependent responses

For LTI systems, the time-invariant response $H_{\mathrm{sTI}}[k]$ is the same irrespective of the test signals. However, again, this is not the case in the real world measurements. We define the LTI response $H_{\mathrm{sLTI}}[k]$ by averaging the responses measured using different test signals. Then, we represent the squared absolute signal-dependent responses $|H_{\mathrm{sSDR}}[k]|^2$ as the average squared absolute value of deviation from the time-invariant response. We use index variable p to identify a member of the set of test signals $\Omega_{\mathrm{P}} = \{\tilde{x}_{\mathrm{s}}^{\{p\}}[n] \mid p = 1, \ldots, P\}$.

$$H_{\mathrm{sLTI}}[k] = \frac{1}{P}\sum_{p=1}^{P} H_{\mathrm{s}}^{(p)}[k] \tag{5}$$

$$|H_{\mathrm{sSDR}}[k]|^2 = \frac{1}{P-1}\sum_{p=1}^{P}|H_{\mathrm{s}}^{(p)}[k] - H_{\mathrm{sLTI}}[k]|^2. \tag{6}$$

Here, $a^{(p)}$ indicates that we used the p-th member of the set Ω_{P} to calculate a. The denominator $P-1$ of Eq. (6) is for unbiased variance.

4. Numerical simulation

We conducted a set of numerical simulations to determine the feasibility of the proposed method. We used white noise for the original test signal and studied the effect of safeguarding by flooring the low-level absolute values.

4.1. Effect of flooring in LTI response

Figure 1 shows the results of simulation using additive white noise with SNR of 40 dB. We adjusted the noise level for the safeguarded signal because flooring increases the signal power. The threshold θ_L value we used here is the average absolute value of the original spectrum. The sampling frequency was 44,100 Hz, and the signal length was 100000 samples.

Figure 2 shows the maximum deviation[5] of the safeguarded gain function for different SNR settings. Flooring significantly reduces the maximum deviations.

Safeguarding by flooring adds a deterministic signal that sounds like noise. The regression analysis of the flooring level $\theta_L|_{\mathrm{dB}}$ (represented in dB) and the power of the deterministic signal (σ_{dB}: also represented in dB) resulted in the following experimental relation:

$$\sigma_{\mathrm{dB}} \sim -10.321 + 1.995\,\theta_L|_{\mathrm{dB}}, \tag{7}$$

where the intercept and the slope indicate that 0 dB flooring adds 10 dB smaller noise and the level decreases two times faster than does the flooring level. For example, the -10 dB flooring level causes SNR to be 30 dB. This relation suggests that safeguarding does not severely damage the quality of the original signal (for example, a music piece).

[4]It is better not to overlap analysis segments.

[5]We used the absolute difference between the true value (0 dB) and the calculated transfer function for deviation.

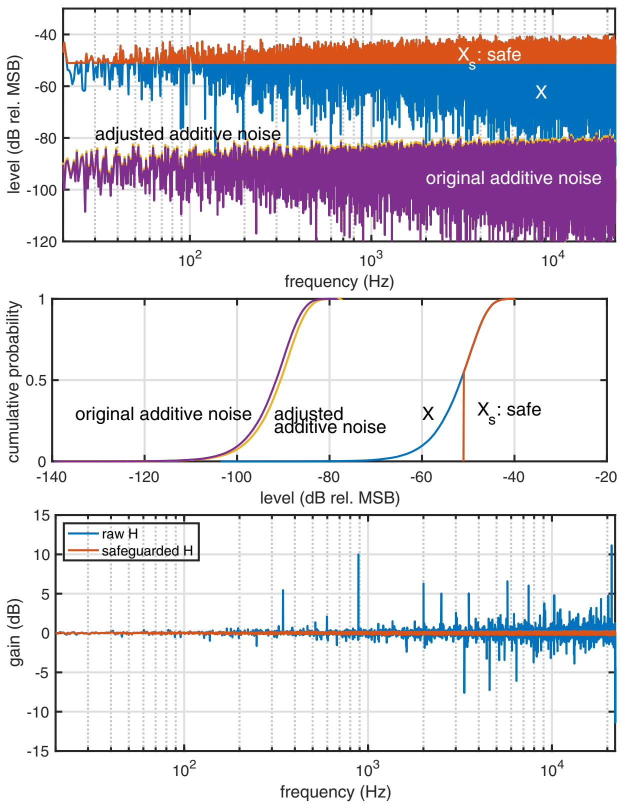

Fig. 1 The top panel shows power spectra of the original white noise, safeguarded noise, and background white noise (original and adjusted). The middle panel shows the level distribution. The bottom panel shows the estimated gains using the original and safeguarded signals.

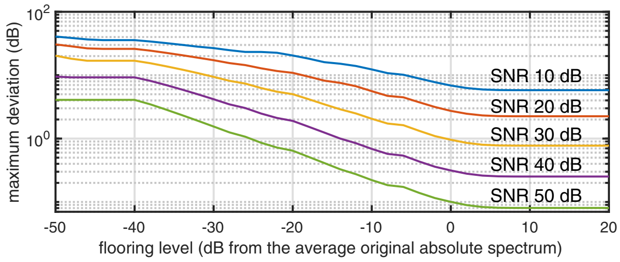

Fig. 2 Effect of flooring on the maximum deviations from the ground truth (0 dB) under different SNR values. A flooring level of -50 dB does not change any frequency bins, whereas that of 20 dB changes all bins.

The estimated gain when using a safeguarded signal still has random peaks and dips. Spectral smoothing is a common practice to elucidate spectral characteristics of acoustical systems. Figure 3 shows the standard deviation of the smoothed gain functions when using one-third-octave-width rectangular smoothing. A comparison with Fig. 2 reveals that spectral smoothing significantly reduces the amount of deviations.

4.2. Random response

Repeated measurements using the same test signal provide the random response estimate using Eq. (4). Figure 4

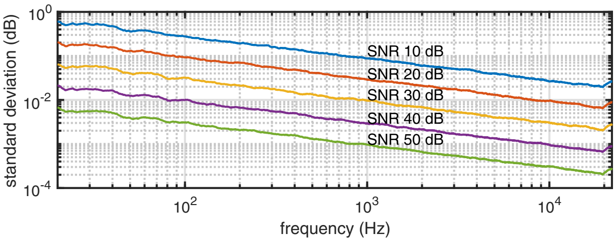

Fig. 3 Standard deviations of the smoothed gain estimated using safeguarded signals. The flooring level is 0 dB and the smoothing width is $1/3$ octave.

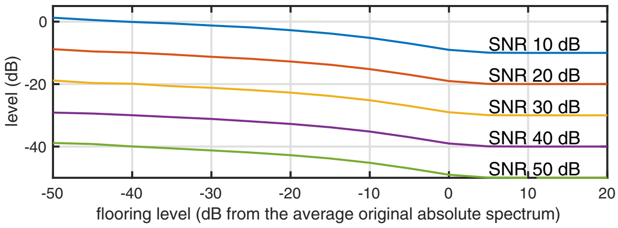

Fig. 4 Estimated random response level dependence on the flooring level.

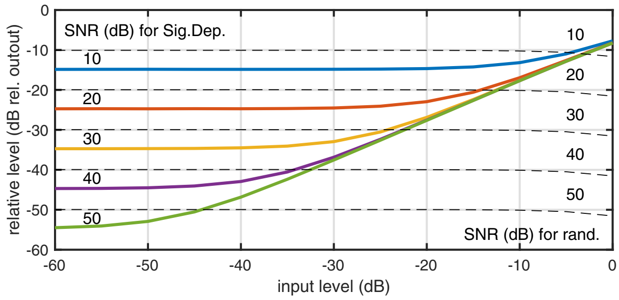

Fig. 5 Input level dependence of the random response and the signal-dependent response ($\alpha = 0.4$).

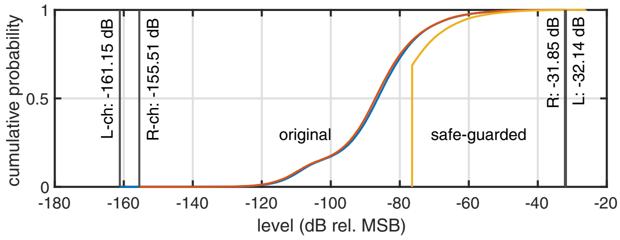

Fig. 6 Absolute value distribution of a loop music segment. We set the threshold θ_s to the average value.

shows the estimated random response level. Note that the estimate at 20 dB flooring provides the correct estimate of the noise level because the test signal has a constant absolute value with a randomized phase; in other words, it is a periodic pseudorandom noise.[6]

4.3. Signal-dependent response

Repeated measurements using the same test signal do not provide signal-dependent deviations such as harmonic distortion and intermodulation distortion caused by nonlinearity. We introduced a sample nonlinearity using

$$y = \frac{1}{\alpha}(\exp(\alpha x) - 1). \tag{8}$$

We used Gaussian white noise as x and added the other Gaussian noise to the output y with the given SNR. We prepared four types of original segment and repeated each segment four times, assuming a piece of typical loop music with the prospective application in mind.

Figure 5 shows the results. Colored lines represent the signal-dependent response level calculated using Eq. (6). The black dashed lines represent the random response level calculated using Eq. (4). The random response levels normalized by the total output levels are virtually constant, which reflects the assigned SNR. The signal-dependent levels decrease as the input level decreases. This decrement saturates at the effective (-3 dB for each doubling of number of repetitions) random response level.

5. Sample measurement using loop music

We conducted acoustic measurements in a connected Japanese room with an area of about 40 m^2. The sound source is a loudspeaker (Fostex FF85WK) with a bass-reflex enclosure (Fostex BK85WB 2). To drive the loudspeaker, we used a power amplifier (Fostex AP20d) connected to an audio interface (PRESONUS STUDIO 2|6). A wide-range omnidirectional condenser microphone (EARTHWORKS M50) connected to the audio interface acquired the reproduced sounds.

We used four loop-music (5 s each) pieces composed for this research. We mixed the stereo track into a monaural track. The sampling frequency was 44,100 Hz.

5.1. Example 1: single test segment

For the first experiment, we positioned the microphone 10 cm in front of the center of the loudspeaker. The sound pressure level (SPL), measured using A-weighting, at the microphone was 95.8 dB. The background noise level was 24 dB. We repeated each segment six times. We used the middle four segments for calculation.[7]

Figure 6 shows the absolute value distribution of the DFT $X[k]$ of the original signal. We floored absolute values at the average absolute value of the DFT.

Figure 7 shows the absolute DFT values of the original and safeguarded signals. The safeguarded signal sounds like the original signal with a slight amount of white noise.

Figure 8 shows the calibrated SPL of the DFT of the acquired loudspeaker output and the background noise. Note that the loudspeaker output in the high-frequency range reflects spectral flooring.

Figure 9 shows the LTI, random and time-varying responses, and the effects of the background noise. We used

[6] CAPRICEP and FVN also provide correct estimates [1,2].

[7] For loop music, it is better to repeat each segment four times and use three segments excerpted after the expected reverberation time plus propagation delay. As long as the length of the segment is exactly identical to that of the signal period, it is not necessary to match the location with the beginning of the segment.

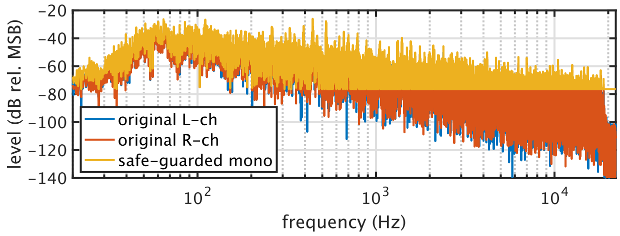

Fig. 7 Absolute value of DFT of the original and safeguarded signals.

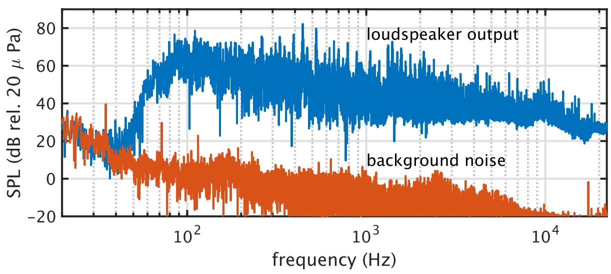

Fig. 8 DFT of the measured segment and background noise segment.

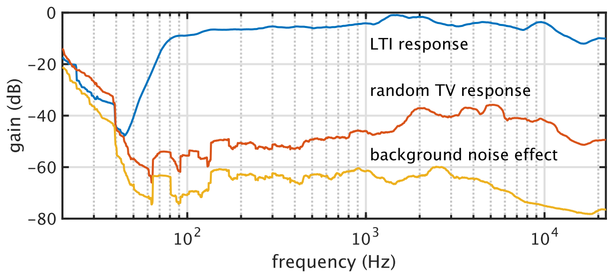

Fig. 9 Separated responses with 1/3 octave smoothing.

1/3 octave smoothing to clarify essential features of the responses. This representation suggests that the random component in the high-frequency region is the result of a high SPL.

5.2. Example 2: multiple test segments

Figure 10 shows the effects of playback levels. The upper plot shows the results of A-weighting SPL at 87.3 dB, and the lower plot at 93.5 dB. The louder playback result[8] shows that the signal-dependent distortion is significantly higher than the random response. The signal-dependent random noise mentioned previously contributes to this difference.

6. Conclusion

We proposed a simple method of measuring the attributes of acoustic systems using arbitrary sounds by safeguarding them. This method enables us to use music to measure acoustic conditions in a room or during concert with an

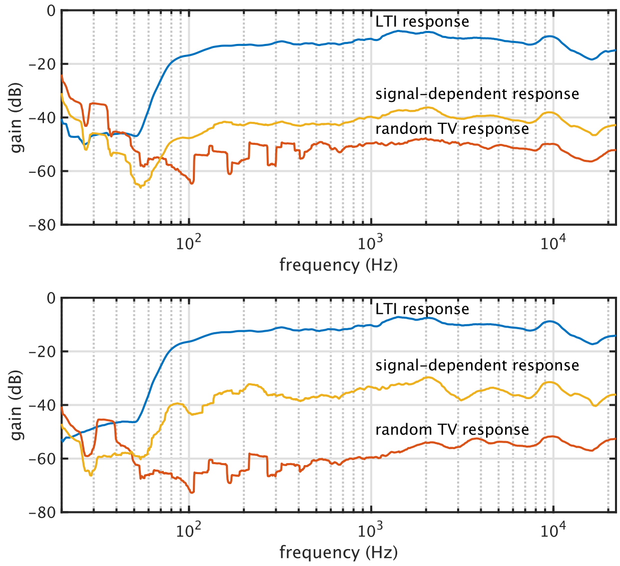

Fig. 10 Measured responses with A-weighting SPLs at 87.3 dB (upper plot) and at 93.5 dB (lower plot).

audience without disrupting their musical experience. In this work, the feasibility of the proposed method was demonstrated. The safeguarding makes it possible to measure the LTI response of the entire frequency range even using an arbitrary signal including a pure tone by adding a slight deterministic noise.

This method has a wide range of applications. For example, for assessing the listening conditions of a classroom, we can use words and phrases. We are planning theoretical and comprehensive investigations of this method and the introduction of frequency-dependent flooring.

Acknowledgement

This work was supported by Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research grant numbers JP21H00497 and JP20H00291. The authors appreciate the discussions and comments of Drs. Ken-Ichi Sakakibara, Tatsuya Kitamura, Mitsunori Mizumachi, and Hideki Banno. The authors greatly appreciate the composition and performance of the loop music by Kazuki Matsumoto.

References

[1] H. Kawahara, K.-I. Sakakibara, M. Mizumachi, M. Morise and H. Banno, "Simultaneous measurement of time-invariant linear and nonlinear, and random and extra responses using frequency domain variant of velvet noise," *Proc. APSIPA ASC 2020*, pp. 174–183 (2020).

[2] H. Kawahara and K. Yatabe, "Cascaded all-pass filters with randomized center frequencies and phase polarity for acoustic and speech measurement and data augmentation," *Proc. ICASSP 2021*, pp. 306–310 (2021).

[8]The LTI gain in Fig. 9 and Fig. 10 are different because they were measured at different times with different sensitivity settings.