

Simultaneous Measurement of Multiple Acoustic Attributes Using Structured Periodic Test Signals Including Music and Other Sound Materials

Hideki Kawahara^{*}, Kohei Yatabe[†], Ken-Ichi Sakakibara[‡], Mitsunori Mizumachi[§], Tatsuya Kitamura[¶],

^{*} Wakayama University, Japan E-mail: kawahara@wakayama-u.ac.jp

[†] Tokyo University of Agriculture and Technology, Japan E-mail: yatabe@go.tuat.ac.jp

[‡] Health Sciences University of Hokkaido, Japan E-mail: kis@hoku-iryu-u.ac.jp

[§] Kyushu Institute of Technology, Japan E-mail: mizumach@ecs.kyutech.ac.jp

[¶] Konan University, Japan E-mail: t-kitamu@konan-u.ac.jp

Abstract—We introduce a general framework for measuring acoustic properties such as liner time-invariant (LTI) response, signal-dependent time-invariant (SDTI) component, and random and time-varying (RTV) component simultaneously using structured periodic test signals. The framework also enables music pieces and other sound materials as test signals by “safeguarding” them by adding slight deterministic “noise.” Measurement using swept-sin, MLS (Maxim Length Sequence), and their variants are special cases of the proposed framework. We implemented interactive and real-time measuring tools based on this framework and made them open-source. Furthermore, we applied this framework to assess pitch extractors objectively.

I. INTRODUCTION

Computing devices today are 10^9 times more powerful than those available a half-century ago [1]. This power makes it possible to process huge amount of speech materials (for example: [2], [3]). This advancement motivated us to establish a framework for making speech materials more usable based on our recent findings [4]–[6].

The speech materials mentioned above do not necessarily fulfill recommended conditions for scientific research [7]. However, the abovementioned conditions are too strict depending on the research purpose. There is room to make speech materials usable by providing tools for assisting data acquisition and retrospective assesment [8].

The framework we introduce in this article provides a solid basis for tools making speech materials reusable. We start with a big picture of the framework, followed by descriptions of constituent algorithms. Then, we introduce acoustic measurement tools followed by applications other than acoustic measurement. Finally, we discuss issues and related works.

The main contribution of this paper is the fundamental reformulation of the underlying algorithms and the complete revision of the infrastructure, computationally efficient implementation based on FFT and inverse FFT. Our previous works consisted of many conceptual confusions and ad hoc and inefficient procedures [4]–[6], [17], [18], [20]–[22]. This paper enables us to renew all applications and tools while maintaining their functionality.

II. MEASUREMENT OF ACOUSTIC ATTRIBUTES

Figure 1 shows a schematic representation of the whole types of acoustic attributes measurement that we will introduce

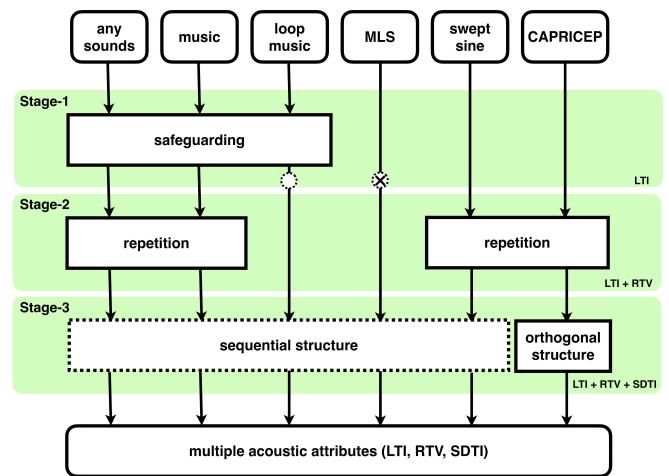


Fig. 1. Acoustic attributes and test signals. (LTI: Linear Time-Invariant, RTV: Random and Time-Varying, and SDTI: Signal Dependent Time-Invariant). The terminal symbol placed at the bottom of Stage-1 on the line from “MLS” represents that results are erroneous. The terminal symbol on the line from “loop music” represents that it destroys the signal design, looping.

in this article. The following three subsections outline this framework followed by sections describing technical details and applications.

A. Linear time-invariant (LTI) response: Stage-1

For linear time invariant (LTI) systems, response to input impulse (impulse response) uniquely determines the target system. In the frequency domain, dividing the Fourier transform of the target system output by the Fourier transform of the input signal provides the transfer function of the target system. The symbol “swept sine” [9] and “MLS” (Maximum Length Sequence [10], [11]) are commonly used test signals. They are members of TSP (Time Stretched Pulse) and have flat power spectrum. The symbol “CAPRICEP” (Cascaded All-Pass filters with Randomized Center frequencies and Phase polarity) is also a new member of TSP we proposed [5].

Spectral division using any other signals also provides the transfer function, in principle. However, their spectrum generally consists of very weak component(s) that makes

spectral division impractical. We introduced “safeguarding” method for making such signal usable for transfer function measurement [6].

B. Random and time-varying component: Stage-2

Two neighboring periods excerpted from the repeatedly concatenated periodic segments are identical. However, periods excerpted from different parts of the acquired signal are not identical because the background noise and sensitivity fluctuations do not have the same periodicity. The difference between observed periods provides background noise and sensitivity fluctuations information. This way, the repetitive presentation of periodic segments makes simultaneous measurement of the LTI response and the disturbance in the measurement.

C. Signal dependent component: Stage-3

Estimated impulse responses using different input signals are identical when the system is strictly LTI. However, estimated impulse responses of an acoustic system in the real world are not identical for different input signals because of non-linearity, even after suppressing the random and time-varying component by averaging many repetitive measurement results. The differences in estimated impulse responses are the second type of disturbing component. The component depends on the used test signals. It is necessary to describe the used test signals.

Repetition (used in Stage-2) is one type of structuring for the test signal. Sequentially aligning different repetitive test signals is the second type of test signal structuring. Using CAPRICEP, we introduce the third type of structuring, simultaneous structuring, by using the orthogonal nature of the Walsh-Hadamard matrix.

III. ALGORITHM

This section describes the underlying algorithms in each stage. The test and the acquired signals are discrete-time signals and share the same sampling clock. (In the appendix, we introduce a solution for handling signals sampled by independent clocks.)

A. LTI response: Stage-1, linear convolution

Let $x[n]$, $y[n]$, and $h[n]$ represent an input signal, an output signal, and the impulse response of a system. The time index n is an integer. Assume that $x[n]$ is non-zero for $0 \leq n \leq N - 1$ and $h[n]$ is non-zero for $0 \leq n \leq M - 1$. Then, the output signal $y[n]$ is non-zero for $0 \leq n \leq M + N - 1$. This setting, usually called as “zero-padding”, is a common practice to calculate linear convolution using DFT-based circular convolution [12]. (See Fig. 2)

By using the discrete Fourier transform $\mathcal{F}[\cdot]$ of length $L > M + N - 1$, the following equation provides the transfer function $H[k]$ of the system. The symbol k represents the index of the discrete frequency.

$$H[k] = \frac{\mathcal{F}[y[n]]}{\mathcal{F}[x[n]]}. \quad (1)$$

Without additive noise, the inverse transform $\mathcal{F}^{-1}[\cdot]$ of the transfer function provides the estimate of the impulse response $h_{\text{est}}[n]$.

$$h_{\text{est}}[n] = \mathcal{F}^{-1}[H[k]]. \quad (2)$$

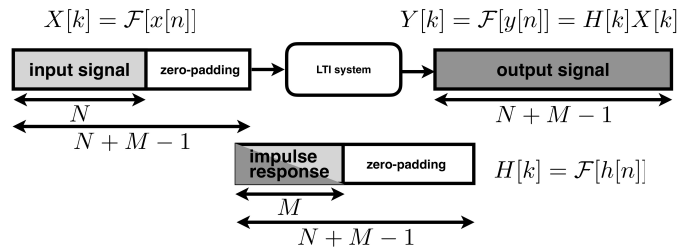


Fig. 2. Convolution of discrete-time signals. Implementation using cyclic convolution with zero-padding.

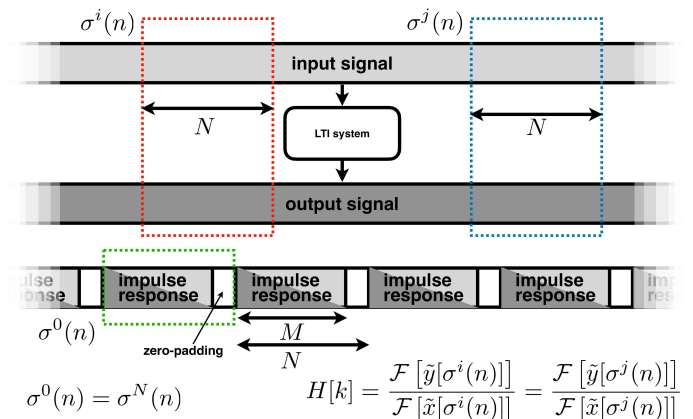


Fig. 3. Convolution of periodic signals. Implementation using cyclic convolution without zero-padding for input and output signals. The same cyclic permutation to input and output does not change the estimated transfer function.

The typical TSP (Time Stretched-Pulse) signal, swept-sine, and the recent member of TSP, unit-CAPRICEP, (usually) have frequency independent gain, such as $|H[k]| = 1$ for all k . Because of the background noise, frequency-dependent gain improves estimation accuracy [13].

1) *Terminal symbols*: Figure 1 has two terminal symbols for “loop music” and “MLS” at the bottom of Stage-1. We placed the symbol because MLS is a periodic sequence and is orthogonal only under cyclic convolution. The separated single cycle of the MLS sequence is not a time-stretched pulse under linear convolution. For loop music, isolating one repeated phrase is usable for calculating the LTI response, although it is not looping and destroys the original design purpose.

2) *Signal safeguarding*: Equation (2) holds for input signals with frequency-dependent spectral values. However, the estimated transfer functions are not usable because small absolute values in the denominator magnify the effects of observation noise. Signal-safeguarding, our proposal, significantly suppresses observation noise effects and makes any sounds appropriate for acoustic measurement [6]. With safeguarding, if necessary, all signals are relevant for acoustic measurement.

B. LTI response, and random and time-varying component: Stage-2, cyclic convolution

Figure 3 summarizes this stage. Zero padding is inefficient. The periodic nature of the discrete Fourier transform removes the zero-padding for input and output in Fig. 2.

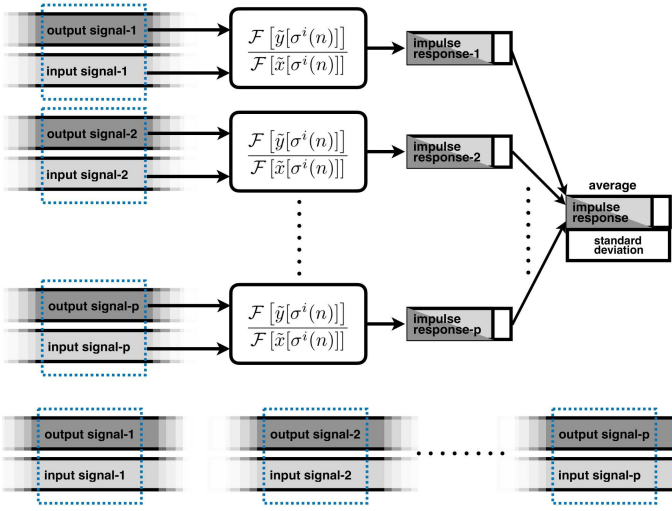


Fig. 4. Serial structuring illustration. Signal-dependent component isolation by averaging and calculating sample standard deviation of estimated impulse responses using different test signals.

Assume that $x[n]$, $y[n]$, and $h[n]$ are periodic and share the same period, L . Then, there is no need for excessive zero-padding for test signals. Using a swept-sine signal in this repetitive presentation, removing zero-valued parts does not cause any problems. The following equation provides the transfer function.

$$H[k] = \frac{\mathcal{F}[\tilde{y}[n]]}{\mathcal{F}[\tilde{x}[n]]}, \quad (3)$$

where $\tilde{x}[n]$ and $\tilde{y}[n]$ represent that the signals are periodic.

Periodic test signal enables repeated measurement. Disturbance in measurement, such as background noise and interfering sounds, does not share the same periodicity with the test signal. Then, it is safe to assume that disturbing sounds are independent at each repetition. Averaging of the estimated impulse responses reduces the estimation variance. This averaging and calculation of observed standard deviation provide an estimate of the random and time-varying component (RTV in Fig. 1).

Periodic test signal has additional merits. Because it is periodic, there is no need to repeat the calculation of the Fourier transform of the test signal. Because it is periodic, there is no need to find the initial position of the analysis segment. Wherever the initial position is, the signal is periodic as far as the length is identical to the period. In short, the following holds.

$$\frac{\mathcal{F}[\tilde{y}[\sigma^i(n)]]}{\mathcal{F}[\tilde{x}[\sigma^i(n)]]} = \frac{\mathcal{F}[\tilde{y}[\sigma^j(n)]]}{\mathcal{F}[\tilde{x}[\sigma^j(n)]]}, \quad (4)$$

where $\sigma^i(n)$ and $\sigma^j(n)$ represent i -th and j -th cyclic permutations of the discrete time sequence n . Figure 3 illustrates these merits.

C. LTI response, and random and time-varying, and signal dependent components: Stage-3, using different test signals

The estimated impulse response is identical when the system is LTI strictly, and no noise exists. However, in actual acous-

tic measurement, impulse responses estimated using different input test signals are different. Averaging impulse responses using many different test signals provides less distorted LTI response. The difference between each impulse response and the averaged one provides an estimate of the magnitude of signal-dependent deviation. Note that we assume each estimated impulse response for a test signal is an average of repetition, and this averaging suppresses random variations.

Figure 4 illustrates one implementation of this strategy, serial structuring. This serial structuring is the only option for test signals other than CAPRICEP.

D. LTI response, and random and time-varying, and signal dependent components: Stage-3, structured test signal

CAPRICEP has another option, orthogonal structuring. Each unit-CAPRICEP has a raised-cosine-shaped power envelope. Periodic allocation of each unit-CAPRICEP with 50% overlap yields a constant power envelope. We use the Walsh-Hadamard matrix B of order 4 to determine the polarity of three different unit-CAPRICEPs.

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} b_{1,1} & \cdots & \cdots & b_{1,4} \\ \vdots & b_{2,2} & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{4,1} & \cdots & \cdots & b_{4,4} \end{bmatrix}. \quad (5)$$

1) *Periodic signal generation:* The following part describes a procedure to make a periodic test signal $\tilde{s}[n]$ from three different unit-CAPRICEPs. The discrete Fourier transform of $\tilde{s}[n]$ yields discrete spectral representation $S[k]$. The following procedure makes the absolute value $|S[k]|$ a constant, that is independent of the discrete frequency k .

Multiplying coefficient $b_{i,j}$ to each unit-CAPRICEP $g_{\text{uC}}^{(q)}[n]$ and applying the overlap-and-add operation with $N/4$ shift for three different unit-CAPRICEPs provide a base unit sequence $u[n]$.

$$u[n] = \sum_{q=1}^3 C_q \sum_{r=1}^4 b_{q,r} g_{\text{uC}}^{(q)} \left[n - \frac{(q-1)N}{4} \right] \quad (6)$$

where the super-script q of $g_{\text{uC}}^{(q)}[n]$ is the identifier of each unit-CAPRICEP. The coefficient C_q value is 1 for $q = 1, 2$, and $\sqrt{2}$ for $q = 3$.

The next step is to make a periodic test signal $\tilde{s}[n]$ from the base unit sequence $u[n]$. Since the width of a unit-CAPRICEP is four to six times wider than the allocation interval $N/4$ (see Appendix), it is necessary to allocate (overlap-and-add) $u[n]$ more than $6 + 1$ times to make a periodic segment. Let P represent the number of repetitions. Considering the condition and for simplicity, we select an even number $P \geq 8$.

The periodic test signal $\tilde{s}[n]$ is an excerpt from the following intermediate signal $s_{\text{tmp}}[n]$.

$$s_{\text{tmp}}[n] = \sum_{p=1}^P u[n - (p-1)N], \quad (7)$$

Excerpting a segment with length N around the center location of $s_{\text{tmp}}[n]$ provides the periodic test signal $\tilde{s}[n]$.

$$\tilde{s}[n] = s_{\text{tmp}}\left[n - \left(\frac{P}{2} - 1\right)N\right], \quad (n = 0, \dots, N-1) \quad (8)$$

2) Simultaneous multiple impulse response measurement:

The following procedure provides three different impulse responses from the output signal $\tilde{y}[n]$. First, similar to the second stage, using the ratio of the discrete Fourier transform of input test signal and output provides an estimate of the impulse response $\hat{h}_L[n]$.

$$\hat{h}_L[n] = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[\tilde{y}[n]]}{\mathcal{F}[\tilde{s}[n]]}\right], \quad (9)$$

where the subscript L of $\hat{h}_L[n]$ represents that it is a longer impulse response obtained from the measurement. The length of $\hat{h}_L[n]$ is L . Although, in actual measurement, other than the initial part, the noise floor (due to background noise and non-linearity, mainly) masks the impulse response corresponds to the LTI part.

The orthogonal structure of $\tilde{s}[n]$ provides three shorter impulse responses by using the virtual target signal $\tilde{v}_S^{(q)}[n]$.

$$\tilde{v}_S^{(q)}[n] = \sum_{r=1}^4 b_{q,r} \delta\left[n, \frac{(q-1)N}{4}\right], \quad (10)$$

where $\delta[i, j]$ is the Kronecker delta and q identifies the corresponding unit-CAPRICEP.

The first $N/4$ elements of the signal defined by the following equation provide a short impulse response $\hat{h}_S^{(q)}[n]$.

$$\hat{h}_S^{(q)}[n] = \mathcal{F}^{-1}\left[\frac{\mathcal{F}[\tilde{v}_S^{(q)}[n]]\mathcal{F}[\tilde{y}[n]]}{\mathcal{F}[\tilde{s}[n]]}\right]. \quad (11)$$

As noted, the initial $N/4$ elements are unique and relevant for impulse response estimates. Deviations from the same initial part of $\hat{h}_L[n]$ represent the signal-dependent component.

Note that $\mathcal{F}[\tilde{v}_S^{(q)}[n]]$ functions as a selector of discrete frequency components. Figure 5 illustrates how it works in the discrete frequency domain. For example, $\hat{h}_S^{(1)}[n]$ selects the blue-colored component, and $\hat{h}_S^{(2)}[n]$ and $\hat{h}_S^{(3)}[n]$ select red-colored and yellow-colored components.

There is a caveat that the procedure above suppresses the signal-dependent component resulting from even-symmetric nonlinearity. It is necessary to estimate impulse responses using the negated test signal $-\tilde{s}[n]$ to detect components caused by even-symmetric nonlinearity.

E. Implementation details

The descriptions mentioned above are simplified summaries. We placed descriptions of technical details in Appendix sections for readability. The discussed details are as follows: a) Optimized weighting shape for reducing artifacts due to truncation of acquired interfering signals. (Appendix A) b) Phase manipulation function that makes better localization of the generated unit-CAPRICEP. (Appendix B) c) Evaluation of signal safeguarding merits. (Appendix C)

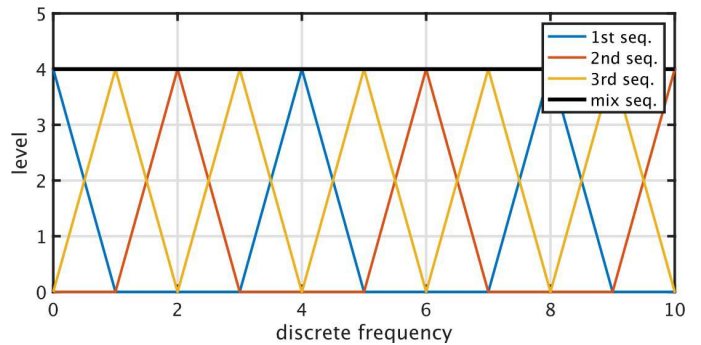


Fig. 5. Discrete spectral levels of the whole test signal $\tilde{s}[n]$ and constituent sequences corresponding first, second, and the third unit-CAPRICEP. The plot shows the initial 11 discrete frequency components.

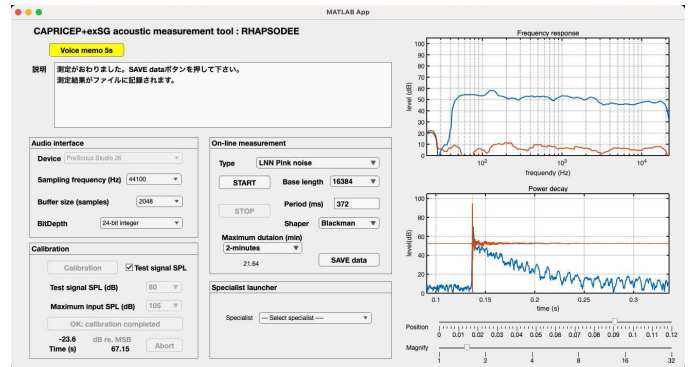


Fig. 6. Snapshot of the GUI of the “Control panel” for acoustic calibration and application launcher [15].

IV. ACOUSTIC MEASUREMENT TOOLS

We developed acoustic measurement tools based on the proposed framework. This section introduces several examples. The tools introduced here are accessible in the first author’s GitHub repositories [14].

A. Control panel for acoustic applications

Figure 6 shows a snapshot of the GUI of the “Control panel” for acoustic calibration and application launcher [15]. The top-left wide field is for showing prompt instructions for users. The left-middle and bottom sub-panels are for setting the audio interface and calibrating input sensitivity. The center sub-panel is for controlling the acoustic measurement. The center-bottom sub-panel is the application launcher.

Figure 7 shows a collection of GUI snapshots of speech-related applications we developed [15]–[18]. Selecting an item from the dropdown menu in the application launcher sub-panel launches the application.

The screenshot of Fig. 6 shows an interactive and real-time measurement of the sound recording environment using a periodic test signal with a modified pink-noise spectral shape. A powered loudspeaker (IKmultimedia iLound Micro Monitor) simulates a talker, and a measuring microphone (Earthworks M50) placed 20 cm in front of the loudspeaker acquires the output sound. The top-right graph shows the smoothed gain (blue line) of the transfer function $H[k]$ and the smoothed

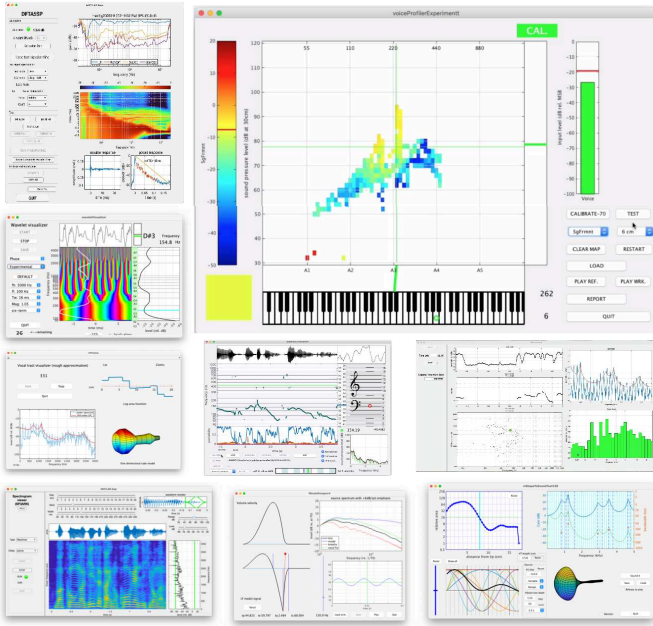


Fig. 7. Snapshots of application GUIs [15]–[18]. This controller launches applications by selecting the dropdown menu in the application launcher sub-panel.

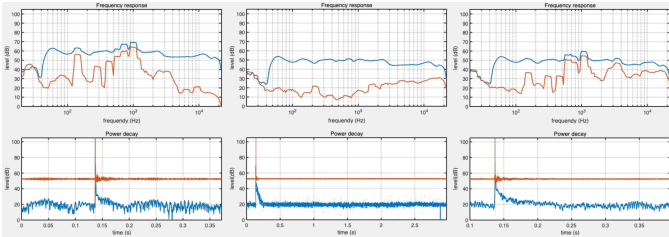


Fig. 8. Snapshots of monitored acoustic conditions. The left shows when the experimenter disturbed the measurement by one’s voice /a/. The center shows the monitored LTI response and background noise using a test sound that is a safeguarded utterance (a Japanese sentence /bakuN ga ginsekai no kougen ni hirogaru/ “A roaring sound spreads across the silvery snow-covered plateau.”) spoken by a Japanese male. These are snapshots excerpted from a demonstration movie.

disturbing component (red line). The following graph shows the impulse response waveform (red line) and the smoothed power of the impulse response represented in dB (blue line).

Figure 8 shows monitored results in different conditions. These snapshots are excerpts from a demonstration movie¹.

The left graph shows when the experimenter disturbs the measurement by producing a loud /a/ voice. The disturbing component in the red line pushes the LTI estimate at several parts. The smoothed impulse response power also shows distortions in low-level parts.

The center graph shows the result using a safeguarded read sentence. The length of the period of the test signal used in this measurement is 2^{17} . Figure 16 in Appendix C shows the spectra of the original and safeguarded signals. The measured LTI gain is virtually identical to the result in Fig. 6. However, the level of disturbing components is higher than Fig. 6. It is

¹<https://youtu.be/-nxD-8hbCv4>

because the power spectral level in the higher frequency region of the safeguarded test signal decreases steeper than that of the pink noise. Note that the noise floor, due to background noise, masks the smoothed power level of the estimated impulse response other than the initial part of the impulse response.

The right graph shows when the experimenter disturbs the measurement by producing a loud /a/ voice, similar to the left graph. Similarly, the disturbing component in the red line pushes the LTI estimate at several parts. There are differences in the shape of the disturbing components in the left and right graphs. The disturbing level is lower in the low-frequency region and higher in the high-frequency region. These differences are due to the spectral difference of the test signals. The pink noise-shaped test signal has higher energy in the high-frequency region than the safeguarded speech sounds.

V. OTHER APPLICATIONS

We developed several tools other than acoustic measurement based on the proposed framework. This section introduces two applications closely related to each other.

The first one is measurement of voice f_0 (fundamental frequency)². This research measures the frequency modulation transfer function using the f_0 frequency modulation by the spectrally shaped test signal made from CAPRICEP as the test signal. Then the f_0 modulation of the produced voice sound as the output.

The second one is an objective measurement of pitch extractors using the f_0 modulation transfer function [22]. In this investigation, we replaced the human with the pitch extractor. The movie that compares 16 pitch extractors using this method is informative³.

VI. DISCUSSION AND RELATED WORKS

The proposed framework does not replace existing acoustic measurement methods. Instead, it provides them additional values and is an efficient computational infrastructure. For example, exponential swept-sine analyzes nonlinearity in a diagnostic way [24]. Our proposed framework provides information on the target system’s behavior handling signals they are designed to use. In this manner, the proposed framework plays a complementing role in acoustic measurement.

Reverberation radius [23] represents the distance where the energy of the direct and the reverberant sounds equals. It is an essential attribute in, for example, classroom acoustics. Our method enables us to measure the reverberation radius and other acoustic attributes using actual teaching materials in a classroom while teaching students. Moreover, using the smartphones of each student, simultaneous measurements of each student’s listening acoustic conditions are possible.

VII. CONCLUSIONS

We introduced a general framework for measuring acoustic properties such as liner time-invariant (LTI) response, signal-dependent component, and random and time-varying component simultaneously using structured periodic test signals. The framework also enables music pieces and other sound materials

²See [19] for the reason why we are using f_0 instead of using F0) response to auditory stimulation [20], [21].

³<https://youtu.be/iXnP1tluVic>

as test signals by “safeguarding” them by adding slight deterministic “noise.” Measurement using swept-sin, MLS (Maximum Length Sequence), and their variants are special cases of the proposed framework. We implemented interactive and real-time measuring tools based on this framework and made them open-source. Furthermore, we applied this framework to assess pitch extractors objectively. The proposed framework is general enough and applicable to many other fields.

ACKNOWLEDGMENT

KAKENHI by JST 21K19794, 21H03468, 21H00497, 21K11957 supported this line of research. The authors appreciate for careful and constructive reviewers’ comments.

REFERENCES

[1] C. E. Leiserson, N. C. Thompson, J. S. Emer, B. C. Kuszmaul, B. W. Lampson, and et al., “There’s plenty of room at the top: What will drive computer performance after moore’s law?” *Science*, vol. 368, no. 6495, p. eaam9744, 2020.

[2] J. Kahn, M. Rivière, W. Zheng, E. Kharitonov, and X. Xu, “Libri-Light: A benchmark for ASR with limited or no supervision,” in *Proc. ICASSP 2020*, May 2020, pp. 7669–7673.

[3] A. Mehrish, N. Majumder, R. Bharadwaj, R. Mihalcea, and S. Poria, “A review of deep learning techniques for speech processing,” *An international journal on information fusion*, p. 101869, Jun. 2023.

[4] 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,” in *Asia-Pac. Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, 2020, pp. 174–183.

[5] H. Kawahara and K. Yatabe, “Cascaded all-pass filters with randomized center frequencies and phase polarity for acoustic and speech measurement and data augmentation,” in *Proc. ICASSP 2021*. [ieeexplore.ieee.org](https://ieeexplore.ieee.org/abstract/document/9415057/), 2021, pp. 306–310. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9415057/>

[6] —, “Safeguarding test signals for acoustic measurement using arbitrary sounds: Measuring impulse response by playing music,” *Acoustical science and technology / edited by the Acoustical Society of Japan*, vol. 43, no. 3, pp. 209–212, 2022. [Online]. Available: <http://dx.doi.org/10.1250/ast.43.209>

[7] R. R. Patel, S. N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyiski, T. Eadie, and et al., “Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing association expert panel to develop a protocol for instrumental assessment of vocal function,” *American journal of speech-language pathology*, vol. 27, no. 3, pp. 887–905, Aug. 2018.

[8] K. Sakakibara and H. Kawahara and M. Mizumachi, “Recommended protocols for acoustic recording of speech data for high reusability,” *Journal of the Acoustical Society of Japan*, vol. 76, no. 6, pp. 343–350, 2020.

[9] N. Aoshima, “Computer-generated pulse signal applied for sound measurement,” *The Journal of the Acoustical Society of America*, vol. 69, no. 5, pp. 1484–1488, May 1981. [Online]. Available: <https://asa.scitation.org/doi/abs/10.1121/1.385782>

[10] J. Borish and J. B. Angell, “An efficient algorithm for measuring the impulse response using pseudorandom noise,” *J. Audio Engineering Society*, vol. 31, no. 7-8, pp. 478–488, 1983. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=4564>

[11] D. D. Rife and J. Vanderkooy, “Transfer-function measurement with maximum-length sequences,” *Journal of the Audio Engineering Society. Audio Engineering Society*, vol. 37, no. 6, pp. 419–444, 1 Jun. 1989. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=6086>

[12] A. V. Oppenheim, R. W. Schaffer, and R. W. Shaffer, *Discrete-time signal processing*, ser. Prentice Hall signal processing series. London, England: Prentice-Hall, Dec. 1989.

[13] Y. Nakahara, Y. Iiyama, Y. Ikeda, and Y. Kaneda, “Shortest impulse response measurement signal that realizes constant normalized noise power in all frequency bands,” *Journal of the Audio Engineering Society. Audio Engineering Society*, vol. 70, no. 1/2, pp. 24–35, 26 Jan. 2021. [Online]. Available: <http://dx.doi.org/10.17743/jaes.2021.0048>

[14] H. Kawahara, “GitHub repository for speech and hearing research/education tools,” 2023, (retrieved 20 June 2023). [Online]. Available: <https://github.com/HidekiKawahara>

[15] H. Kawahara, K.-I. Sakakibara, and K. Yatabe, “Singing voice range profiling toolbox with real-time interaction and its application to make recording data reusable,” in *Proceedings of the Stockholm Music Acoustics Conference 2023*, S. D’Amario, A. Friberg, and S. Ternström, Eds., 14 Jun. 2023, pp. 160–166. [Online]. Available: <https://smcnetwork.org/smc2023/>

[16] H. Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, T. Toda, and T. Irino, “A new cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis,” in *Proc. Interspeech 2017*, Stockholm, Aug. 2017, pp. 1358–1362. [Online]. Available: <http://dx.doi.org/10.21437/Interspeech.2017-15>

[17] H. Kawahara, K.-I. Sakakibara, E. Haneishi, and K. Hagiwara, “Real-time and interactive tools for vocal training based on an analytic signal with a cosine series envelope,” in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2019, pp. 907–910.

[18] H. Kawahara, K. Yatabe, K.-I. Sakakibara, M. Mizumachi, M. Morise, H. Banno, and T. Irino, “Interactive and Real-Time acoustic measurement tools for speech data acquisition and presentation: Application of an extended member of time stretched pulses,” in *Show and Tell Proc. Interspeech 2021*, 2021, pp. 4853–4854. [Online]. Available: https://www.isca-speech.org/archive/pdfs/interspeech_2021/kawahara21b_interspeech2021.pdf

[19] I. R. Titze, R. J. Baken, K. W. Bozeman, S. Granqvist, N. Henrich, C. T. Herbst, D. M. Howard, E. J. Hunter, D. Kaelin, R. D. Kent, J. Kreiman, M. Kob, A. Löfqvist, S. McCoy, D. G. Miller, H. Noé, R. C. Scherer, J. R. Smith, B. H. Story, J. G. Švec, S. Ternström, and J. Wolfe, “Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization,” *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 3005–3007, 2015. [Online]. Available: <http://dx.doi.org/10.1121/1.4919349>

[20] H. Kawahara, T. Matsui, K. Yatabe, K.-I. Sakakibara, M. Tsuzaki, M. Morise, and T. Irino, “Mixture of Orthogonal Sequences Made from Extended Time-Stretched Pulses Enables Measurement of Involuntary Voice Fundamental Frequency Response to Pitch Perturbation,” in *Proc. Interspeech 2021*, 2021, pp. 3206–3210.

[21] H. Kawahara, T. Matsui, K. Yatabe, K. I. Sakakibara, M. Tsuzaki, M. Morise, and T. Irino, “Implementation of interactive tools for investigating fundamental frequency response of voiced sounds to auditory stimulation,” in *Proc. APSIPA ASC*, Tokyo, Dec. 2021, pp. 897–903.

[22] H. Kawahara, K. Yatabe, K.-I. Sakakibara, T. Kitamura, H. Banno, and M. Morise, “An objective test tool for pitch extractors’ response attributes,” in *Proc. Interspeech 2022*, 2022, pp. 659–663. [Online]. Available: <http://dx.doi.org/10.21437/Interspeech.2022-800>

[23] M. Mijić and D. Mašović, “Reverberation radius in real rooms,” *Telfor journal*, vol. 2, no. 2, pp. 86–91, 2010. [Online]. Available: https://journal.telfor.rs/Published/Vol2No2/Vol2No2_A6.pdf

[24] A. Novak, L. Simon, and P. Lotton, “Analysis, synthesis, and classification of nonlinear systems using synchronized Swept-Sine method for audio effects,” *EURASIP journal on advances in signal processing*, vol. 2010, no. 1, pp. 1–8, 20 Jul. 2010. [Online]. Available: <http://dx.doi.org/10.1155/2010/793816>

APPENDIX

A. Reducing truncation artifacts

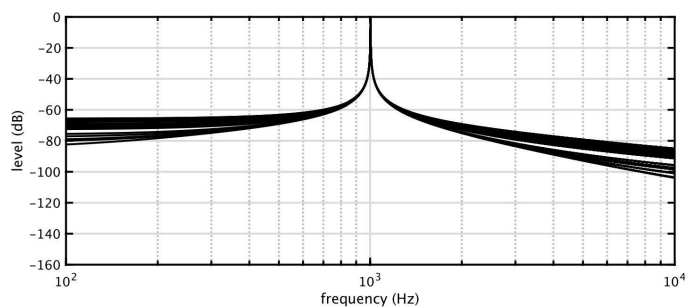


Fig. 9. Power spectra of truncated signal without any shaping the both ends.

Figure 9 shows an example of signal truncation asynchronous to the signal. This example uses a sinusoid having a frequency $1000 + \pi$ Hz. As shown in the figure, truncation produces spectral spread.

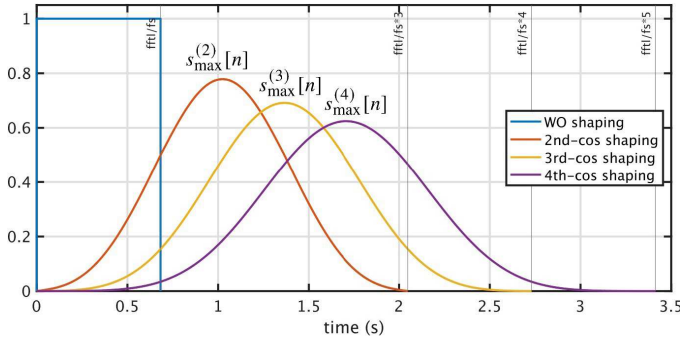


Fig. 10. Periodicity preservation tests of shaping functions. Plot shows the weighting function shapes.

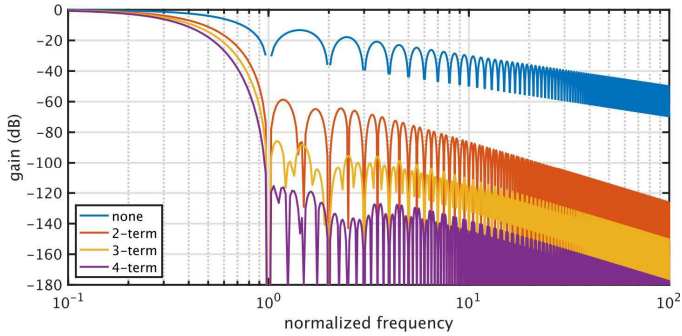


Fig. 11. Optimized weighting functions made from several cosine series convolved with a rectangular function with the repetition period length. They are numerically optimized to minimize the maximum side-lobe level.

Using more than one cycle of the periodic signal and a weighting function reduces this issue. The weighting function has to be a constant value one when wrapped. Figure 10 shows such weighting functions. They are cosine-based window functions convolved with a rectangular function having the period width. The annotation $s_{\max}^{(k)}[n]$ indicates that the convolved cosine series has k -terms. The coefficients are numerically optimized to minimize the maximum side-lobe level.

Figure 11 shows their frequency responses.

Figure 12 shows the same asynchronous signal truncated using the optimized weighting function $s_{\max}^{(4)}[n]$. Note that weighting effectively suppressed the spectral spreads.

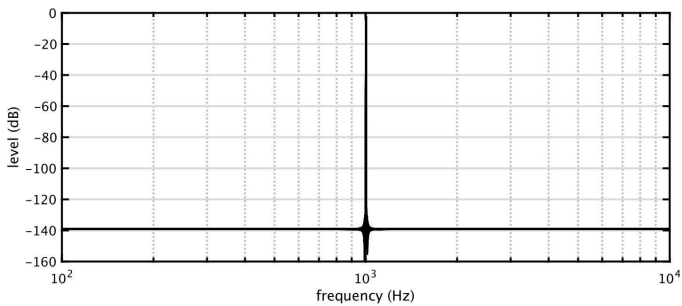


Fig. 12. Power spectra of truncated signal with optimized shaping weight.

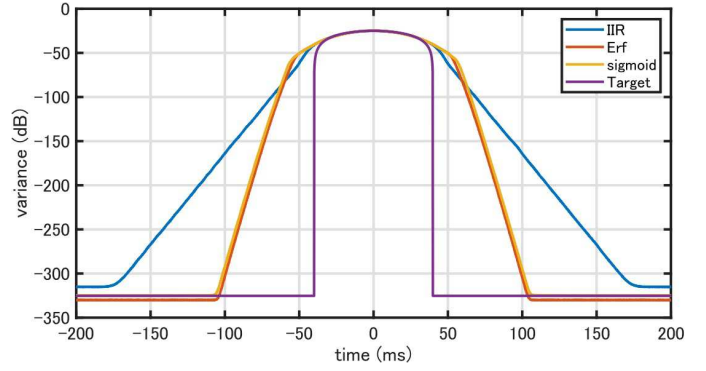


Fig. 13. Designed outline shapes using a raised cosine shape as the target. The vertical axis represents the variance in logarithmic (dB) scale.

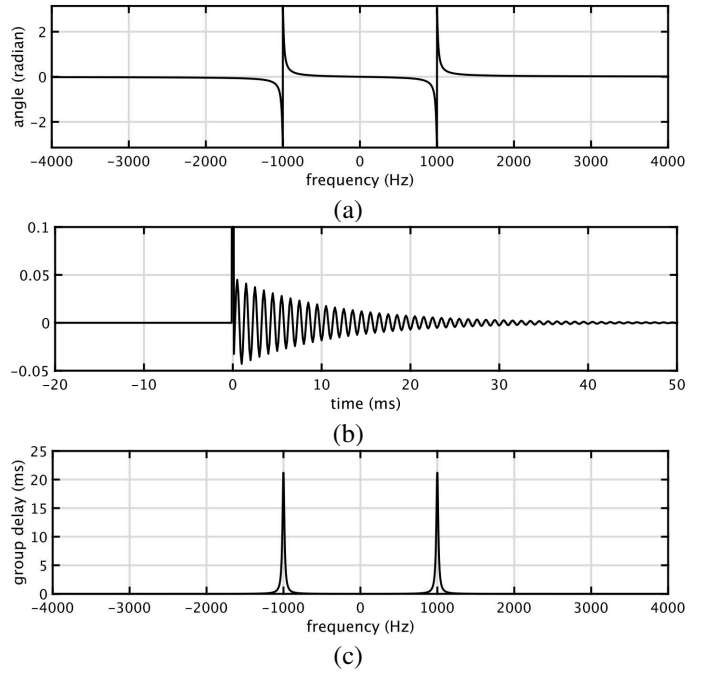


Fig. 14. Attributes of IIR-type all-pass filter. (a) phase-frequency response, (b) impulse response, and (c) group delay response.

B. Phase manipulation function

Let us start with the result. Figure 13 shows the RMS average of the unit-CAPRICEP designed using a phase manipulation function of the original proposal [5] (IIR) and the error function, integrated Gaussian function (Erf) and its approximating function sigmoid (sigmoid).

Figure 14 shows an example of relations between phase manipulation, impulse response, and the group delay of the initially proposed phase function [5]. It is the phase of an IIR all-pass filter. This manipulation example and the following examples use phase manipulation at one point, 1000 Hz. The actual unit-CAPRICEP of 0.25 s effective width consists of about 7000 phase manipulation points.

Figure 15 shows impulse responses using the error function and the sigmoid. The bottom plot compares IIR, Erf, and Sigmoid's power decay. The decay of Erf is the steepest, and we

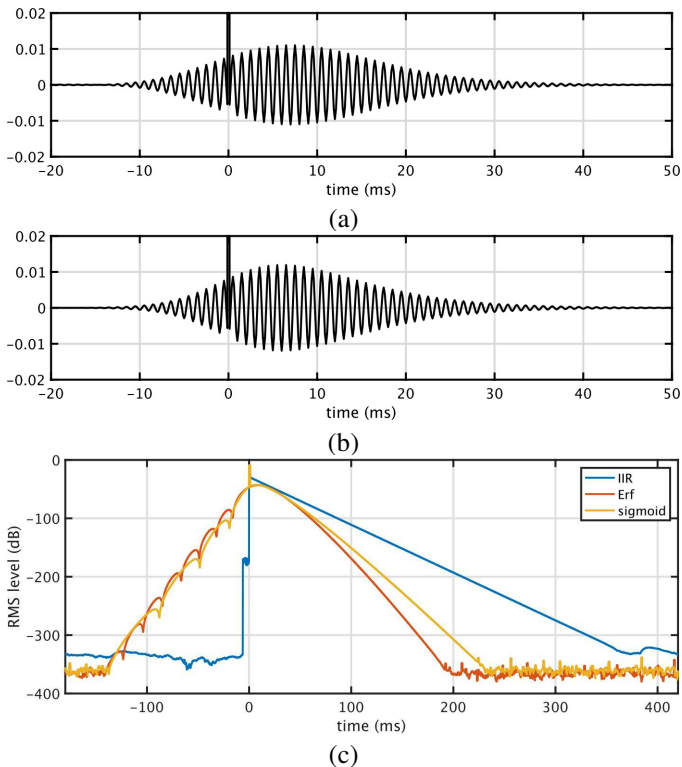


Fig. 15. Impulse response of the all-pass filters made from error function (a) and the sigmoid (b). The plot (c) shows smoothed RMS levels (in dB) of the impulse response of the IIR (original CAPRICEP [5]), the error function (Erf), and the sigmoid (sigmoid).

selected Erf to design unit-CAPRICEP in our implementation.

C. Safeguarded signal and its merits

This is a revised excerpt form [6]. Let $x[n]$ be a periodic discrete-time signal with a period L . 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 DFT (Discrete Fourier transform) 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 DFT, where k , ($k = 0, \dots, L - 1$), 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 have larger value than a threshold⁴. We use the following equation to derive the DFT $X_s[m]$ of the safeguarded signal $\tilde{x}_s[n]$.

$$X_s[k] = \begin{cases} \frac{\theta_L[k]X[k]}{|X[k]|} & \text{for } 0 < |X[k]| < \theta_L[k] \\ X[k] & \theta_L \leq |X[k]| \end{cases}, \quad (12)$$

⁴In actual implementation we use frequency dependent threshold $\theta_L[k]$ using power spectra of the original signal and the background noise. We also set the minimum level and low frequency limit for $\theta_L[k]$.

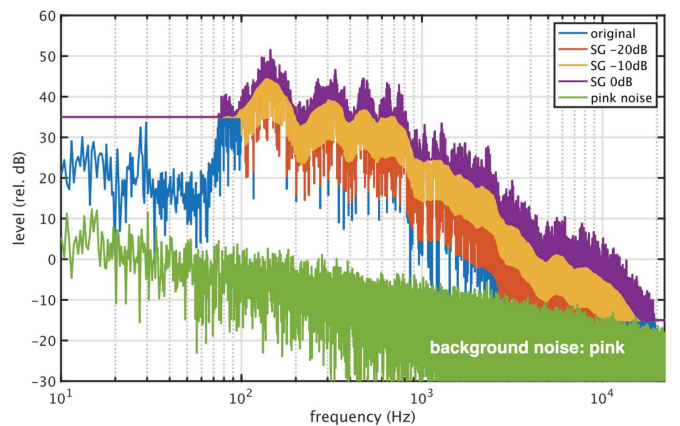


Fig. 16. Absolute values of the original (a Japanese sentence /bakuon ga ginsekai no kougē ni hirogaru/ spoken by a male speaker) and safeguarded speech samples. The signal period is 2^{17} samples at 44100 Hz. Frequency-dependent thresholding uses a smoothed power spectrum with 1/3 octave width as a reference. The light green line represents the background pink noise for testing noise tolerance.

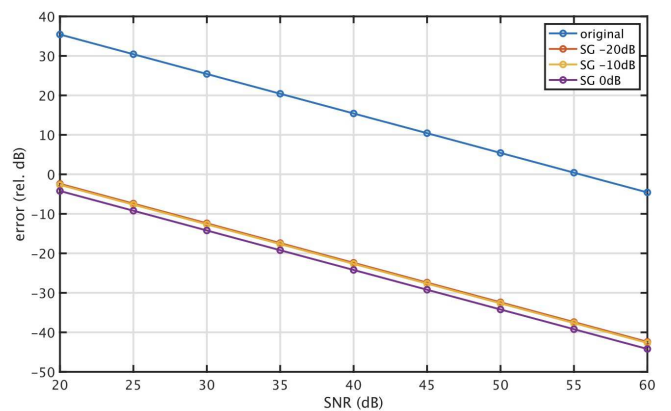


Fig. 17. Estimation error of the impulse responses using the original and the safeguarded signals for the test signal. For noise tolerance test, we added a background pink noise as shown in Fig. 16.

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

$$H_s[k] = \frac{Y_s[k]}{X_s[k]}, \quad (13)$$

where $Y_s[k]$ represents the DFT of the output of the target system for 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 concatenating it as many times as required.

Figure 16 shows example of safeguarding. The level represents the absolute values of discrete Fourier transform of a whole sentence length samples (the original and safeguarded ones). Figure 17 illustrates the merit of safeguarding.