INTERSPEECH 2021: Show & Tell Contribution
30 August – 3 September, 2021, Brno, Czechia

Interactive and real-time acoustic measurement tools for speech data acquisition and presentation: Application of an extended member of time stretched pulses

Hideki Kawahara[1], *Kohei Yatabe*[2], *Ken-Ichi Sakakibara*[3], *Mitsunori Mizumachi*[4], *Masanori Morise*[5], *Hideki Banno*[6], *Toshio Irino*[1]

[1]Wakayama University, Wakayama, 640-8510 Japan
[2]Waseda University, Tokyo, 169-8555 Japan
[3]Health Science University of Hokkaido, Hokkaido, 061-0293 Japan
[4]Kyushu Institute of Technology, Kitakyushu 804-8550 Japan
[5]Meiji University, Tokyo, 164-8525 Japan
[6]Meijo University, Nagoya, 468-8502 Japan

kawahara@wakayama-u.ac.jp, k.yatabe@asagi.waseda.jp, kis@hoku-iryo-u.ac.jp, mizumach@ecs.kyutech.ac.jp, mmorise@meiji.ac.jp, banno@meijo-u.ac.jp, irino@wakayama-u.ac.jp

Abstract

Objective measurements of speech data acquisition and presentation processes are crucial for assuring reproducibility and reusability of experimental results and acquired materials. We introduce setting and measurement examples of those conditions using an interactive and real-time acoustic measurement tool based on an extended time-stretched pulse. We also introduce supporting tools.

Index Terms: speech data acquisition, speech presentation, acoustic measurement, time-stretched pulse

1. Introduction

Proper preparation and recording of conditions when acquiring speech materials are crucial for assuring reusability of the materials [1, 2]. It is also crucial to measure and document speech presentation conditions for making the results reproducible. We introduce an interactive and real-time tool for acoustic measurement and its supporting tools. The tools implement simultaneous measurement procedure of acoustic properties [3] upgraded by introducing an infrastructure based on a new member of time-stretched pulses [4]. We made the tools open-sourced [5][1] and prepared introduction materials including videos [6].

2. Interactive and real-time tool

Figure 1 shows a snapshot of the GUI of the acoustic measurement tool [7]. The left panel consists of control buttons. The right panel consists of real-time data visualization and manipulation tools. The log of operation and acquired data are saved with unique names based on the time-stamp.

2.1. Control panel (left panel)

The bottom three buttons are for "START," "STOP," and "QUIT," the real-time operation and terminate the tool. The red button above provides five-second recordings of time-stamped and voiced comments. The middle sub-panel is for calibrating the input sensitivity. The top sub-panel is for detailed measurements of multiple acoustic properties.

[1]CAPRICEP repository consists of them. CAPRICEP stands for Cascaded All-Pass filters with RandomIzed CEnter frequencies and Phase polarity.

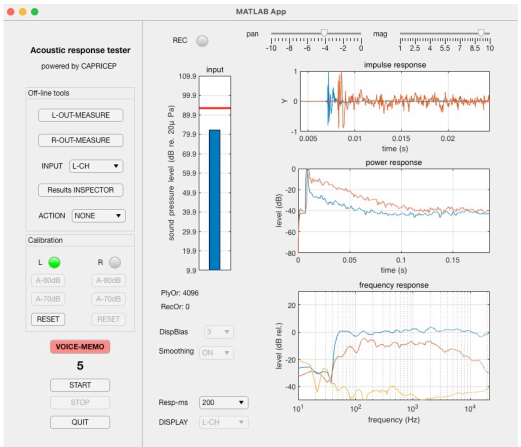

Figure 1: *Graphical User Interface (GUI) of an interactive and real-time acoustic measurement tool based on CAPRICEP [4].*

2.2. Real-time panel (right panel)

The left bar graph monitors input levels (the blue bar indicates the RMS value, and the red line indicates the peak value). The top-right graph shows impulse responses of two acoustic paths (from two loudspeakers to one microphone). The middle-right graph shows the smoothed power responses of two acoustic paths. The decay rate provides a brief grasp of room reverberation. The bottom-right graph shows the gain frequency characteristics of two paths and disturbing components (background noise, errors due to movement, and other random causes).

2.3. Detailed measurement of multiple attributes

Figure 1 shows an example of the detailed measurement result of an acoustic path. This example uses a 192 kHz sampling frequency using an omnidirectional measurement microphone (EARTHWORKS M50) which has a flat response up to 50 kHz. The loudspeaker is a one-inch full-range (AURA SOUND NSW1-205-8A) installed in a closed enclosure. The distance between the microphone and the loudspeaker is 5 cm.

The top graph shows the frequency response of the linear time-invariant responses (thin gray line: raw response, and thick black line: 1/6 octave smoothed response), the nonlinear time-invariant response (yellow line), random and time-varying response (violet line), and the background noise measured while

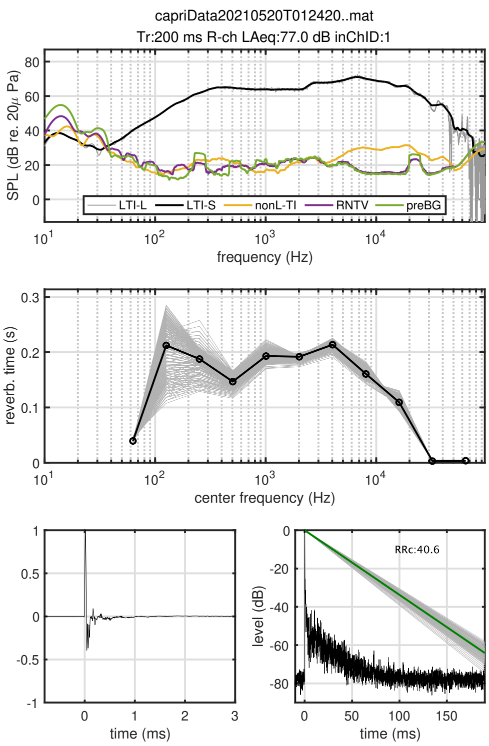

Figure 2: *Examples of simultaneous measurement of acoustic parameters [7].*

Figure 3: *Measurement examples.*

preceding one second (green line).

The middle graph shows the octave-band reverberation time. The gray lines show results using several internal parameter settings. The thick black line indicates the median of them.

The bottom-left graph shows the impulse response of the initial 3 ms. The bottom-right graph shows the power response and the critical ratio RRC. The RRC value times the acoustic path length provides the distance where the power due to direct path and reverberation are the same. The large RRC value and low background level assure relevant voice acquisition quality.

2.4. Supporting tools

Figure 3 shows measurement settings prepared for video presentation. These examples use the measurement microphone as the reference and a large diaphragm microphone (RODE NT2-A) and a dynamic type microphone (SHURE SM48S-LC) as the target. One of the supporting tools provides comparisons of several setting results.

Figure 4 shows the effects of using a sound shield with omnidirectional pattern (red line) and cardioid pattern (black line) in voice acquisition. The omnidirectional microphone suffers

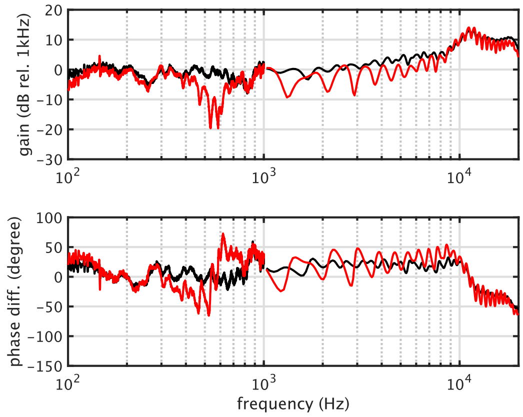

Figure 4: *Gain and phase response dependency of sound shield. Black lines represent cardioid pattern results and red lines represent omni-directional results [7].*

significantly. The results indicate that using a microphone with an omnidirectional pattern with a sound shield is harmful.

3. Conclusions

Show and Tell presentation of these tools is beneficial for researchers to acquire more reusable speech materials and make their research results more reproducible. The underlying infrastructure of these tools, CAPRICEP, has further applications in speech science.

4. Acknowledgements

This work was supported by JSPS (Japan Society for the Promotion of Science) Grants-in-Aid for Scientific Research Grant Numbers JP18K00147, JP18K10708, JP19K21618, and 21H04900.

5. References

[1] R. R. Patel, S. N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J. G. Švec, and R. Hillman, "Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function," *Am. J. Speech-Lang. Pathol.*, vol. 27, no. 3, pp. 887–905, 2018.

[2] J. G. Švec and S. Granqvist, "Guidelines for selecting microphones for human voice production research," *Am. J. Speech-Lang. Pathol.*, vol. 19, no. 4, pp. 356–368, 2010.

[3] 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.

[4] 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. ICASSP2021*, 2021, pp. 306–310.

[5] H. Kawahara, "GitHub repository for speech and hearing research/education tools," 2021, (retrieved 20 May 2021). [Online]. Available: https://github.com/HidekiKawahara

[6] Search "HidekiKawahara" on YouTube.

[7] H. Kawahara, K. Yatabe, K.-I. Sakakibara, M. Mizumachi, M. Morise, H. Banno, and T. Irino, "Tools and practice for supporting recommended protocol for acoustic recording of speech data for high usability – Application of an extended time stretched pulse based on a cascaded all-pass filters with randomized center frequencies and phase polarities," *IEICE Tech. Report*, vol. IEICE-121, no. 66, pp. 1–6, 2021, (in Japanese).