Development of an easily-usable smartphone application for recording instrumental sounds

Takanori Horibe and Masanori Morise *

Meiji University
{cs222034, mmorise}@meiji.ac.jp

Abstract. We have studied the automatic performance skill evaluation in the instrumental sound based on only the recorded sound. Since many instrumental sounds are essential for statistical analysis, a tool for effectively collecting instrumental sounds is helpful. This paper introduces an easily-usable smartphone application that users can record their performance with a single tap operation. This application has several functions to appropriately reject the insufficient result based on simple acoustical analysis.

Keywords: recording application, instrumental sound, acoustical analysis

1 Introduction

Instrumental sound analysis has been carried out from several aspects [1–3], and the recording tool is essential to collect the sounds. To record the instrumental sounds with high quality, the researchers often employ a recording engineer and use a quiet environment, such as a soundproof room and recording studio. On the other hand, if the background noise in the environment does not affect the acoustic analysis, it is unnecessary to record the sound with such quality. In such a case, it is reasonable to record the instrumental sounds by each user with a smartphone application.

In this study, we developed a smartphone application for recording instrumental sounds with a fundamental frequency that the user can easily record the sound. Since the player and the recording operator use this application for recording, the user includes both of them in this study. This application has several functions to automatically reject insufficient results based on acoustic analysis. The concept of this application is that non-expert can record instrumental sounds with a simple operation.

* We thank Dr. T. Oku for providing the instrumental sound required for making Fig. 2. This work was supported by JSPS KAKENHI Grant Number JP21H04900.

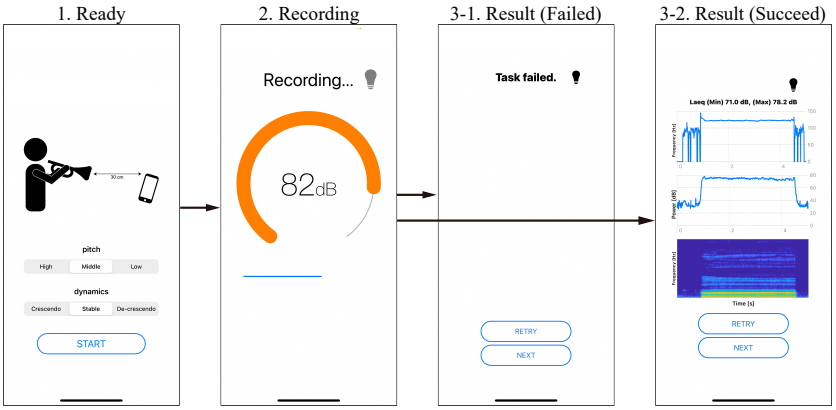

Fig. 1. Screenshots of the implemented application. The user can record the instrumental sounds with a single tap of the START button, and the insufficient result is automatically rejected.

2 Concepts and implementation

2.1 Concepts of the application

It is assumed that users cannot always record their performances in soundproof rooms or recording studios and may record in general rooms. Since background noise is contained in the instrumental sound recorded in such an environment, it is desirable to guarantee enough sound pressure level (SPL) of instrumental sounds. Based on the above, we attempted to implement an interface and functions that meet the following two points.

- When this application is launched, the recommended distance between the smartphone and the instrument is provided as a guide for recording at the sufficient SPL. The recording can be completed with a single button tap operation.
- A simple analysis is performed immediately after the recording to automatically detect errors such as missed performances, clipping, and inappropriate recording, for example, far distance from the smartphone.

2.2 Procedure for recording

Fig. 1 shows the procedures for recording by using the implemented application. After setting the recording condition by the left panel of Fig. 1, the user can record the instrumental sounds with only a single tap. We explain the detailed procedure as follows.

First, the user places a smartphone in an appropriate position and launches this application. On the *1. Ready* panel, the user selects the pitch and volume of the instrumental sound from the "pitch" and "dynamics" boxes and starts recording by the START button. As shown in *2. Recording* panel, the application records instrument sound for 5 seconds after the 3-second countdown. During recording, the user plays a

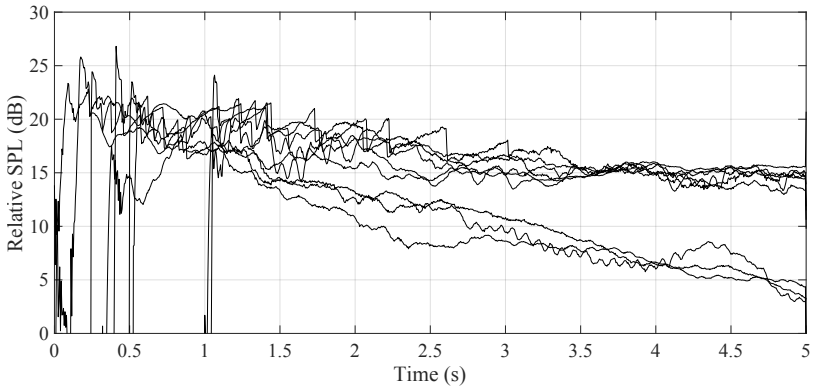

Fig. 2. Time sequences of the relative SPL of ten trumpet sounds played as the decrescendo.

long tone with monitoring the equivalent continuous A-weighted SPL. Clipping is automatically detected when the maximum absolute amplitude of the instrumental sound exceeds 0.95. If clipping, the icon in the upper right corner lights up, and the application skips the following step. If not, a simple acoustic analysis is carried out to reject the insufficient result.

When the result does not meet the required quality, it moves to *3-1. Result (Failed)* panel and instructs re-recording. If it does, it moves to *3-2. Result (Succeed)* panel. It feeds back the fundamental frequency contour, power based on equivalent continuous A-weighted SPL, and spectrogram of the instrumental sound to the user. The user can select whether to Accept or Reject by referring to these results; if Accept, the user can return to *1. Ready* panel by the NEXT button, and if Reject, the user can move to *2. Recording* panel by RETRY button and start over from the countdown. The accepted result is automatically saved with a filename based on the selected conditions.

2.3 Implementation

This application was developed with Swift. The instrumental sounds are recorded using the AVAudioRecorder [4] of the AVFAudio Framework. The sampling conditions are 48 kHz, 16 bits with PCM format. The equivalent continuous A-weighted SPL displayed on *2. Recording* panel was calculated with a frame length of 200 ms.

An acoustic analysis to determine the validity of the recording results is as follows. First, the fundamental frequency and spectral envelope are calculated from the recorded instrumental sounds using WORLD [5]. The spectral envelope is summed for each frame to obtain the relative SPL. The validity of the recording results is judged based on the relative SPL, and the insufficient results are rejected.

This application can reject the sound with too high SPL by detecting the clipping. On the other hand, provided the distance between the microphone and the instrument is too far, the instrumental sound is not likely to be recorded with sufficient SPL. To solve this problem, we test-recorded ten decrescendo trumpet sounds with this application with the sufficient condition and confirmed the relative SPL. Fig. 2 shows the relative

SPLs of all results. The horizontal and vertical axes show the time and relative SPL, respectively. According to this result that the relative SPLs were included from 3 to 27 dB, we calculated a median value from the relative SPL in all frames identified as voiced section. When the median value is in the range of 3–27 dB, the result is accepted. This calculation enables to reject where the SPL of the instrumental sound is too low.

3 Discussion

The user can use this application from recording to analysis with a single tap operation. The application can reject the insufficient result. The clipping detection rejects the recorded sound with too high SPL. The identification by the threshold also rejects the recorded sound with too low SPL. Since this function does not require an environment such as a soundproof room, the user can record their performance in a general room.

Since our research target is the automatic performance skill evaluation by recorded instrumental sounds, the next step requires the evaluation of the application by recording many kinds of sounds with many players. We can evaluate the application in two aspects; One is whether the insufficient result is appropriately rejected. The other is whether the sufficient result is appropriately accepted. The evaluation also includes the usability evaluation of whether the user can easily use this application.

4 Conclusion

In this study, we implemented a smartphone application to record instrumental sounds and confirmed that users can record, analyze, and save a file with a single tap operation. This application has several functions to obtain only reliable sounds. Informal tests confirmed that this application could automatically reject the insufficient sound.

We will statistically evaluate the performance of implemented functions by many recording results. After confirmation of the adequacy of them, we will collect the instrumental sounds by various kinds of users. Developing acoustic features related to the performance skill by using the recording sounds is also an important future work.

References

1. N. Kimura, K. Shiro, Y. Takakura, H. Nakamura, and J. Rekimoto: SonoSpace: Visual Feedback of Timbre with Unsupervised Learning, 28th ACM International Conference on Multimedia, pp. 367–374 (2020)
2. T. Knight, F. Upham, and I. Fujinaga: The potential for automatic assessment of trumpet tone quality, 12th International Society for Music Information Retrieval Conference, pp. 573–578 (2011)
3. A. Lee, S. Furuya, M. Morise, P. Iltis and E. Altenmüller: Quantification of instability of tone production in embouchure dystonia, Parkinsonism & Related Disorders, vol. 20, no. 11, pp. 1161–1164 (2014)
4. "AVAudioRecorder", `https://developer.apple.com/documentation/avfaudio/avaudiorecorder`
5. M. Morise, F. Yokomori, and K. Ozawa: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, IEICE transactions on information and systems, vol. E99-D, no. 7, pp. 1877–1884 (2016)