A quantitative evaluation of a musical performance support system utilizing a musical sophistication test battery

Yasumasa Yamaguchi[1], Taku Kawada[2], Toru Nagahama[3] and Tatsuya Horita[3]

[1] Sendai University
[2] Sendai Shirayuri Gakuen Elementary School
[3]Graduate School of Information Sciences, Tohoku University
ys-yamaguchi@sendai-u.ac.jp

Abstract. This article discusses the effects of a pitch feedback system integrated into Google Glass, called the MVP (Musical pitch Visualization Perception) support system, on the musical performance of wind instrumentalists. The study adopted the Goldsmith Musical Sophistication Index (Gold-MSI) to discuss the contribution of the MVP support system to the improvement of musical performance. The Gold-MSI is a popular tool in the field of music research that measures musical sophistication based on observable behaviors. The study reports the effects of the MVP support system from a quantitative standpoint, and the results show that the system had a positive impact on the participants' pitch accuracy.

Keywords: ICT, Performance support, Pitch Feedback, Performance evaluation, Quantitative analysis

1 Introduction

The intonation of instrumentalists has been extensively discussed in the literature [1], with correct intonation being viewed as particularly important for novice instrumentalists when performing in an ensemble. Effective methods for improving intonation have been widely studied in music education research [2]. In recent years, many researchers have shown an interest in real-time pitch feedback systems, owing to the development of Information and Communication Technologies (ICT). For instance, Wang et al. [3] investigated the potential of real-time feedback for violinists. This study, however, will focus on wind instrumentalists who perform in a concert band or orchestra.

Instrumentalists must pay attention not only to their intonation but also to other crucial aspects of performance, such as sheet music, rhythm, dynamics, fingering, tempo, expression, ensemble, and conductor cues[4]. Information provided by the conductor is as essential as auditory information. However, during performances, some novice and student instrumentalists tend to rely on their tuner, which is typically placed on their music stand. This habit may cause fixation of gaze, narrowing of visual field, and poor posture. Therefore, Yamaguchi et al. [5,6]proposed a pitch feedback system integrated into Google Glass as a Musical pitch Visualization Perception (MVP) support system to solve the problems mentioned above.

Yamaguchi et al. [5,6] argue that the MVP support system has advantages over the conventional tuner in that it provides flexibility to the instrumentalist's physical and visual angle and reduces cognitive load during performance, according to usability tests and interview surveys. However, these qualitative methods have not escaped criticism, such as subjectivity and generalizability. Furthermore, these kinds of investigations do not fully represent the contribution of the musical performance system. The difficulty of quantitative analysis is an overall problem regarding musical performance research, however, evaluation of musical skill and ability have been conveniently defined by years of experience of musical activity. Yet these ideas should be dealt with as a multifaceted and complex concept. Thus, some of the aforementioned, convenient definitions are unworthy of trust. Controlling for musical sophistication, including skill and ability, is a major problem which still exists in the literature.

In this research, we report the effects of the MVP support system from a quantitative standpoint. To discuss the contribution of the MVP support system to the improvement of musical performance, we adopted the Goldsmith Musical Sophistication Index (Gold-MSI) [7, 15]. It is promising that the level of musical sophistication, such as skill and experience, will greatly affect the effectiveness of musical performance support systems.

2 MVP support system

The MVP support system has been developed for wind instrumentalists. Although, the main concept was a pitch feedback system for them, it would also become a performance support system by utilizing Google Glass. The system can be defined as a glass-type tuner for instrumentalists and shows promise regarding the reduction of physical burden and the improvement of gaze flexibility and cognitive load during musical performances when compared to using a conventional tuner on a music stand [5,6].

The feedback system utilized a three-tier scale rating system of "correct", "higher", or "lower" compared to the correct pitch, with the participant receiving real-time feedback through color indicators on the display. The correct pitch range was defined as the target, expressed in Hertz, ±1% [5,6]. The Glass Enterprise edition 2 by Google was used in the study. The reliability of the Google Glass system as a musical tuner was verified by a professional musician and the first author, who has experience conducting and training concert bands (for a detailed explanation, see [5]). The system was developed with four key standpoints: timeliness, ease of understanding, recordability, and stability [5]. To achieve this goal, the system utilized the ml5.js library [8], run in TensorFlow, which includes the PitchDetection package that implements the deep-learning-based CREPE algorithm [9,10]. In this study, we also used the Glass Enterprise edition 2 by Google to send tonal pitch feedback to participants in real time. We also recorded the performance data and pitch estimate data in a CSV file, with the evaluation of the performance pitch stored in a separate column [5]. The schematic drawing of the MVP support system is shown in Figure 1.

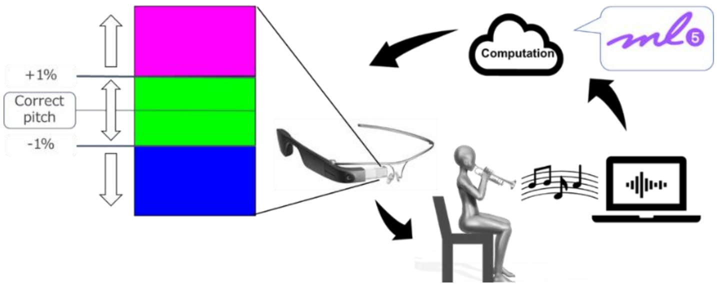

Fig. 1. The schematic of the MVP support system

3 Goldsmith Musical Sophistication Index

The Gold-MSI is a popular tool in the field of music research [7]. The term "musical sophistication" was introduced by Müllensiefen and colleagues as a more inclusive and neutral alternative to terms such as "talent" and "aptitude," and is based on observable behaviors. The Gold-MSI is composed of 38 self-report questions divided into five subscales: Active Engagement, Perceptual Abilities, Musical Training, Singing Abilities, and Emotions. These questions were carefully selected from a pool of 153 statements extracted from previous studies. The Gold-MSI has been validated by a large number of primarily English-speaking participants and has demonstrated good internal reliability for each subscale as well as the overall sophistication index, high test-retest reliability, and reliable correlation with a variety of objective listening ability tests.

The Gold-MSI has been widely used since its inception in 2014 and has been translated into various languages including Traditional Chinese, Simplified Chinese, Portuguese, German, and French [11-14]. These translations have shown high internal consistency and test-retest reliability, and the validation data collected from these studies indicate a good fit with the bifactor model structure proposed by the original Gold-MSI. These findings suggest that the structure and set of questions used to measure musical sophistication by the Gold-MSI are applicable to other cultures and languages.

The development of batteries for assessing musical abilities in Japan has resulted in the creation of several tests, including the Onken Musical Aptitude Test for Young Children [16], and the New Musical Aptitude Test [17]. Most of these tests are designed to assess the musical abilities of children. While there are few non-Japanese standardized tests that have been translated into Japanese, some, such as the Bentley Measure of Musical Abilities [18], have been translated and are available for use. However, there is currently no validated Japanese version of the Gold-MSI, which is widely used to assess musical sophistication.

Therefore, Sadakata et al. [15] translated the Gold-MSI into Japanese (Gold-MSI-J) and, after validating the translation with 689 Japanese speakers, it was found that the internal consistency and test-retest reliability were excellent. Furthermore, the confirmatory factor analysis showed that the bifactor model structure proposed by the original study of Gold-MSI is reasonably maintained in the data.

4 Method

4.1 Participants

We conducted the experiment which consisted of 22 student participants with an average age of 22.14 years (SD=1.39). Students were selected from national universities' concert band and orchestra clubs. Their musical ability as wind instrumentalists was confirmed to be at an intermediate or advanced level by the Japanese version of Goldsmith Musical Sophistication Index (Gold-MSI-J) [15]. Their musical training factor score in Gold-MSI-J was 3 or more. The participants were 4 clarinet players, 4 trumpet players, 4 flute players, 3 French horn players, 2 saxophone players, 2 trombone players, 1 oboe player, and 1 bassoon player.

4.2 Procedure

The study comprised two distinct sections, namely a performance task section and a questionnaire and interview section. In the performance task section, the participants were instructed to perform the B-flat equal temperament major scale in long-tone while synchronizing with a metronome presented on the screen in front of them, set to a beat per minute (BPM) of 60. This long-tone scale task is the fundamental performance for wind instrumentalists in Japanese concert band societies. To gauge the efficacy of the Google Glass tuner system, a commercially available conventional tuner (YAMAHA TDM-70) which was placed on a music stand, was employed as a baseline. The order effects were counterbalanced for the performance task section. In both the baseline and Google Glass settings, the participants were advised to focus on their tempo and intonation while performing the task. There was no repetition and the duration of the experiment was approximately 30 minutes including explanation and warming up. After the performance section, we conducted a questionnaire survey regarding the musical sophistication using Gold-MSI-J. Finally, we ensured that all measurements were made using the same microphone, computer, browser, network environment and Google Glass system.

5 Results

Figure 2 shows the boxplots regarding the scores of the Gold-MSI-J subscales. Compared to the results of Sadakata et al. [15], these scores clearly show that musical sophistication of the participants is much higher than laypeople.

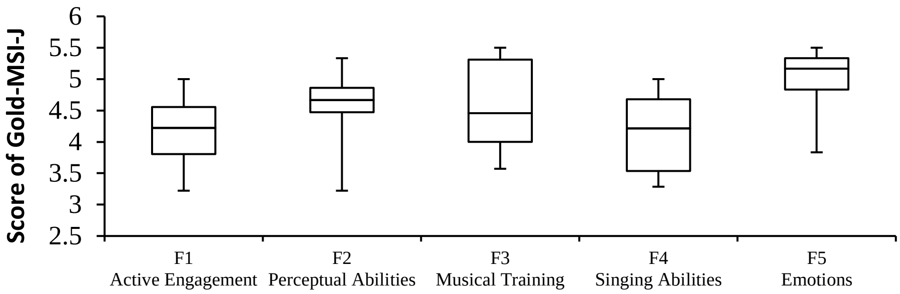

Fig. 2. The boxplots regarding the results of Gold-MSI-J

Table 1. List of participants and pitch accuracy data

ID	Instrument	Conventional tuner	Google Glass
1	Clarinet	78.52%	81.87%
2	French Horn	62.26%	91.93%
3	Euphonium	73.61%	86.70%
4	Trombone	75.65%	76.91%
5	Oboe	94.52%	96.58%
6	Flute	98.95%	97.75%
7	Trumpet	91.79%	95.16%
8	French Horn	84.67%	82.78%
9	Clarinet	54.88%	62.58%
10	French Horn	82.36%	92.86%
11	Flute	87.49%	96.18%
12	Trombone	54.66%	72.18%
13	Trumpet	85.56%	89.90%
14	Trumpet	93.68%	87.45%
15	Saxophone	71.54%	82.78%
16	Flute	95.32%	91.34%
17	Flute	97.63%	95.61%
18	Saxophone	87.59%	88.24%
19	Clarinet	94.70%	94.22%
20	Bassoon	95.46%	96.07%
21	Clarinet	94.37%	97.65%
22	Trumpet	76.15%	85.28%

The median score of each factor in Gold-MSI-J was used as the threshold to divide participants into low and high groups for each factor, which were then used as between-participant factors. The tuner use condition and the Google Glass use condition were defined as within-participant factors. Two-way repeated measures analysis of variance, for the effects of musical sophistication and the tuner system condition, was conducted on the pitch accuracy during the performance. Pitch accuracy was estimated by the recorded data of the MVP support system and transformed to angle data before analysis because it was ratio data by arcsine transformation. Table 1 shows the instrument of each participant and the pitch accuracy under both conditions. The accuracy rate was calculated by dividing the length of time the performed pitch was within the correct pitch range by the total length of the performance.

The results show that there was no significant main effect between participant groups for the Active Engagement factor ($F(1,20)=0.31$, *ns.*, $\eta_p^2=0.02$), and only the main effect of tuner system factors was significant ($F(1,20)=9.68$, $p<.01$, $\eta_p^2=0.33$). There was no significant interaction between Active Engagement factor and tuner system factors ($F(1,20)=2.68$, *ns.*, $\eta_p^2=0.19$). For the Perceptual Abilities factor, there was no significant main effect between participant groups ($F(1,20)=1.90$, *ns.*, $\eta_p^2=0.09$), and only the main effect of tuner system factors was significant ($F(1,20)=5.27$, $p<.05$, $\eta_p^2=0.21$). There was no significant interaction between Perceptual Abilities factor and tuner system factors ($F(1,20)=3.27$, *ns.*, $\eta_p^2=0.14$). For the Musical Training factor, both the main effect between participant groups ($F(1,20)=6.04$, $p<.05$, $\eta_p^2=0.23$) and the main effect of tuner system factors ($F(1,20)=9.96$, $p<.01$, $\eta_p^2=0.31$) were significant. There was no significant interaction between Musical Training factor and tuner system factors ($F(1,20)=2.44$, *ns.*, $\eta_p^2=0.11$). However, this factor can be hypothesized that it has an effect of the performance, we conducted exploratory testing of simple main effects. The result showed that pitch accuracy increased significantly only when the Google Glass tuner was used in the low group of between-participants factor ($F(1,20)=8.92$, $p<.01$, $\eta_p^2=0.47$). For the Singing Abilities factor, there was no significant main effect between participant groups ($F(1,20)=1.26$, *ns.*, $\eta_p^2=0.06$), and only the main effect of tuner system factors was significant (F $F(1,20)=7.43$, $p<.05$, $\eta_p^2=0.27$). There was no significant interaction between Singing Abilities factor and tuner system factors ($F(1,20)=2.68$, *ns.*, $\eta_p^2=0.03$). For the Emotions factor, there was no significant main effect between participant groups ($F(1,20)=0.32$, *ns.*, $\eta_p^2=0.02$), and only the main effect of tuner system factors was significant ($F(1,20)=6.32$, $p<.05$, $\eta_p^2=0.24$). There was no significant interaction between Emotions factor and tuner system factors ($F(1,20)=2.26$, *ns.*, $\eta_p^2=0.10$).

For all analyses, the main effect of tuner system factors was significantly higher for Google Glass tuner conditions than for tuner conditions in terms of pitch accuracy. The main effect between participant groups for the Musical Training factor showed that the high group had significantly higher pitch accuracy than the low group. For this factor, the result of the test of simple main effects showed that the Google Glass conditions significantly improved pitch accuracy only in the low group.

6 Discussion

When using the Google Glass tuner, the accuracy of participant's pitch showed good values. Regarding the relationship with Gold-MSI-J, participants were divided into high and low groups based on their factor scores, and for all factors of Gold-MSI-J, the main effect of the tuner system was significant. On the other hand, the main effect between participant groups was significant only for the Musical Training factor. Furthermore, as a significant difference in pitch accuracy was observed only in the low group in the simple main effect test, it is suggested that the Google Glass tuners may have a strong effect as a performance support system for individuals with relatively low levels of musical training among instrument players. In the high musical training group, pitch accuracy was sufficiently high even under tuner use conditions, and there was no significant difference between the two conditions. However, the sample size is relatively small, so the statistical analysis was conducted with the aim of obtaining an overview of the data. Thus, we will have to conduct large-scale experiments for valid discussions, for example, the influence different instruments have on the results and the usablitity of the system. It will be also necessary to rely on qualitative analysis or subjective evaluations of participants, as described below, to understand how this system contributes to such individuals.

The results of this study show the effect of the MVP support system from each factor of musical sophistication by Gold-MSI-J. Compared with existing qualitative surveys, this method allowed a valid investigation regarding the individual performance and musical sophistication from multifaced standpoints. It should be noted that the present study also emphasized the effectiveness of the MVP support system. Further work in this area is underway to develop an effective pitch feedback system that is suitable to each instrumentalist. It is hoped that the outcome of this study will be of use for future empirical research regarding musical performance study.

References

1. Madsen, C. K. , & Geringer, J. M. Preferences for trumpet tone quality versus intonation. *Bulletin of the Council for Research in Music Education,* 46, 13–22 (1976).
2. Elliott, C. A. Effect of Vocalization on the Sense of Pitch of Beginning Band Class Students. *Journal of Research in Music Education*, 22(2), 120–128. (1974).
3. Wang, J. H., Wang, S. A., Chen, W. C., Chang, K. N., & Chen, H. Y. *Real-time pitch training system for violin learners.* In 2012 IEEE International Conference on Multimedia and Expo Workshops. 163-168) IEEE. (2012, July).
4. Drake, C., & Palmer, C. Skill acquisition in music performance: relations between planning and temporal control. *Cognition*, 74(1), 1-32. (2000).
5. Yamaguchi, Y., Kawada, T., Nagahama, T., & Horita, T. A Pilot Study of the MVP Support System using Google Glass. In EdMedia+ Innovate Learning.17-28. Association for the Advancement of Computing in Education (AACE). (2022, June).
6. Yamaguchi, Y., Kawada, T., Nagahama, T., & Horita, T. Development and Evaluation of a Musical Instrument Performance Support System Using Smart Glasses: from the Subjective

Evaluation of Form, Gaze, and Performance. *Japan journal of educational technology*,46. 185-188. (2023).

7. Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. The musicality of non-musicians: An index for assessing musical sophistication in the general population. In J. Snyder (Ed.), PLoS ONE (Vol. 9, Issue 2, p. e89642). https://doi.org/10.1371/journal.pone.0089642. (2014).
8. NYU. ITP *"ml5js· Friendly Machine Learning for the Web."* ml5js website. Accessed March 16, 2022. https://ml5js.org/, last accessed 2023/5/4.
9. Kim, J. W., Salamon, J., Li, P., & Bello, J. P. *Crepe: A convolutional representation for pitch estimation.* In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 161-165). IEEE. (2018).
10. Mathieu, B., Essid, S., Fillon, T., Prado, J., & Richard, G. *Yaafe, an easy to use and efficient audio feature extraction softwarem* In Ismir. (2010).
11. Degrave P., Dedonder J. A French translation of the Goldsmiths Musical Sophistication Index, an instrument to assess self-reported musical skills, abilities and behaviours. Journal of New Music Research, 48(2), 138–144. https://doi.org/10.1080/09298215.2018.1499779 (2019).
12. Lima C. F., Correia A. I., Müllensiefen D., Castro S. L. Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. Psychology of Music, 48(3), 376–388. (2020).
13. Lin H. R., Kopiez R., Müllensiefen D., Wolf A. The Chinese version of the Gold-MSI: Adaptation and validation of an inventory for the measurement of musical sophistication in a Taiwanese sample. Musicae Scientiae, 25(2), 226–251. (2021).
14. Schaal N., Bauer A.-K., Müllensiefen D. Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe. [The Gold-MSI: Replication and validation of a survey instrument for measuring musical sophistication with a German sample.] Musicae Scientiae, 18(4), 423–447. 10.1177/1029864914541851. (2014).
15. Sadakata, M., Yamaguchi, Y., Ohsawa, C., Matsubara, M., Terasawa, H., von Schnehen, A., Müllensiefen, D., & Sekiyama, K. The Japanese translation of the Gold-MSI: Adaptation and validation of the self-report questionnaire of musical sophistication. Musicae Scientiae, 0(0). https://doi.org/10.1177/10298649221110089. (2022).
16. Ongaku Shinri Kenkyu Sho. Onken-shiki Youji no ongaku tokusei tesuto [Onken's Musical Aptitudes Test for Young Children]. Nihon Bunka Kgakusha Co. Ltd. (1969).
17. Ogawa Y., Murao T., Mang E. H. S. Developing a music aptitude test for school children in Asia [Paper presentation], 10th International Conference for Music Perception and Cognition, Hokkaido, Japan. (2008).
18. Furuichi H., Umemoto T. Bentley ongaku nouryoku tesuto no hyoujun ka [The translation and standardization of the Bentley Measure of Musical Abilities]. Journal of the Musicological Society of Japan, 21, 65–77. (1975).
19. Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
20. Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
21. LNCS Homepage, http://www.springer.com/lncs, last accessed 2016/11/21.