Effective Textual Feedback in Musical Performance Education: A Quantitative Analysis Across Oboe, Piano, and Guitar

Rina Kagawa[1], Nami Iino[2,3], Hideaki Takeda[2], and Masaki Matsubara[1] *

[1] University of Tsukuba
[2] National Institute of Informatics
[3] RIKEN Center for Advanced Intelligence Project
kagawa-r@md.tsukuba.ac.jp

Abstract. Despite the importance of feedback in musical performance education, there is a lack of quantitative and cross-instrumental examination on what feedback is effective for students. This study collected recordings of performances by students on three instruments (oboe, piano, and guitar) and gathered written feedback from multiple teachers for each performance. Quantitative analysis revealed that the usefulness of feedback varied significantly among teachers, independent of musical instruments, compared to pieces or students. We then conducted multilevel modeling based on hierarchy among teachers for each instrument, and found that the number of sentences *giving objective information* significantly contributed to the usefulness of feedback. Our findings have high generalizability and can be applicable to face-to-face lessons. The collected recordings and written feedback have been published, and can provide valuable resources for music educators seeking to improve their teaching practices.

Keywords: Database, Music Education, Verbal Information

1 Introduction

Traditionally, people are taught to play musical instruments face-to-face. However, remote lessons can enable the provision of instruction in remote areas, flexible scheduling, reduced travel, security and cost, and can enhance teachers' and students' creative learning and critical thinking by reducing time and distance between teacher-student or teachers [2, 45]. To increase the value of remote lessons, a set of remote lesson modules

* This study was partially supported by JST-Mirai Program Grant Number JPMJMI19G8, JSPS KAKENHI Grant Number JP19K19347, and Kayamori Foundation of Informational Science Advancement. We would like to thank all the performers and teachers who participated in the data collection of this study. We would also thank to those who helped us with data annotation and evaluation.

(e.g., the Swing project[4]) and an online platform for music distance learning education and practice (e.g., Intermusic project[5]) has developed.

During the COVID-19 pandemic, the demand for remote lessons increased significantly [1, 19]. In fact, online lessons and asynchronous lessons in which critiques are given to recordings were provided at some music colleges, and students and teachers deeply recognized the benefits and shortcomings of remote lessons [27]. In the post-pandemic era, remote lessons will lead to a new group of students joining the music education environment, which will bring diversity to music education and facilitate its development. We believe that music performance instruction will take place in a variety of learning formats, including online, real-world, and hybrid formats.

In traditional face-to-face lessons, the teachers use non-verbal information, such as singing melodies and making gestures, and verbal information, such as their knowledge of the piece [9, 13, 24]. However, in remote lessons, it is difficult to convey detailed body movements and high-quality sound due to the low resolution of video and audio. Therefore, the quality of verbal feedback must be improved to continue to increase the value of remote lessons. However, what kind of content should be included in verbal feedback for students across musical instruments has not been clarified.

Therefore, this study asks, what kind of verbal information related to musical performance is useful for music students? In order to present generalized findings, this study collects and integratively analyzes musical performance data and verbal feedback data with respect to performances on multiple musical instruments.

We have been conducting this study since 2020 and have already published data for the oboe [26]. In this study, we collected additional data for two more musical instruments and analyzed them in an integrated manner.

The contributions of this paper are as follows.

- Musical performance data and corresponding verbal instruction data have been collected and published for three musical instruments.
- The usefulness of verbal feedback on performance was found to depend on the teacher, not the piece or the student, and this finding was independent of the musical instruments.
- The contents of verbal feedback that significantly affected the usefulness of the feedback independent of the musical instruments were clarified.

2 Related Work

2.1 Music Database for Research

Several datasets have been published as music knowledge resources [32]. They adopt various perspectives, including performance recordings [14, 16], metadata (genre, composer, lyrics, fingering, music analysis, etc. [17, 18, 31, 35, 38]), musical scores [12, 22, 23, 42], other multimodal information [25, 43], emotions [6, 7, 46], and students' interpretations [20, 21, 30, 33]. To the best of our knowledge, none have focused on teaching behavior for musical performance. There are also several datasets for music listen-

[4] https://aec-music.eu/project/swing-2018-2021-erasmus-strategic-partnership/

[5] https://aec-music.eu/project/intermusic-2017-2020-erasmus-strategic-partnership/

ing events [34, 40], but these are datasets gathered from online music services such as Last.fm [6] and they are not intended for pieces for student to learn performance.

Compared to these databases, this study is novel as it collected a dataset that allows for the evaluation of the relationship between the content of the verbal information in the instructions and the musical performances.

2.2 Effects of Teaching Behavior on Musical Performance Education

The effects of teaching behavior on musical performance education have been widely studied within the field of music education. Prior studies compared teacher levels [15], analyzed time allocation [5], compared [44] and categorized [36, 37] verbal and non-verbal information, and examined teacher-student interaction [11]. These studies all depended upon the transcription of speech in interactive instruction. There are also studies on supporting the learning of musical performance by presenting nonverbal information [39, 41]. Unlike these studies, our study focused on verbal feedback, which is more applicable to asynchronous education.

One study compared verbal and non-verbal instruction [8], and another study evaluated and summarized the usefulness of instruction [10]. Both were based on five or fewer performances. In contrast, we have conducted a large-scale experiment to clarify the relationship between verbal information and its usefulness.

3 Materials

In our previous studies [28, 29], we collected the performance recordings of the oboe and corresponding textual feedback and published them as CROCUS (CRitique dOCUmentS of musical performance) dataset.[7] In this study, we collected similar data for the piano as a keyboard instrument and the guitar as a string instrument and published them.[8] Then, each sentence of textual feedback was annotated to indicate what was described, and the perceived utility of each piece of textual feedback was evaluated. [9]

All procedures have been approved by the ethical review board of University of Tsukuba, Senzoku Gakuen College of Music, and Kunitachi College of Music.

3.1 Recording

An overview of materials is presented in Table 1 and 2. We selected the pieces shown in Table 3 considering the balance of difficulty, style, form, and era.
piano: We used the recording data collected and published in the previous study [20]
oboe: We used the recording data collected and published in the previous study [28, 29]. Each student played all 10 pieces in a less reverberant and less noisy environment at home, about one meter away from the recording device (Roland R-07).

[6] https://www.last.fm/
[7] https://masaki-cb.github.io/crocus/
[8] **piano:** https://zenodo.org/record/7753365, **guitar:** https://zenodo.org/record/7778923
[9] For oboe, the procedure of questionnaire survey regarding the usefulness and annotation of the textual feedback was the reprint of [29].

Table 1. Overview of Materials

	Recording			
Instrument	$N_{student}$	Level of student	N_{piece}	recording
piano	4	professional players	10	home
oboe	9	music college student	10	home
guitar	12	Students who have participated in national or international competitions	7	home

Table 2. Overview of Materials

	Textual feedback		
Instrument	$N_{teacher}$	Level of teachers	$N_{textual feedback}$
piano	24	professional teachers	144
oboe	12	music college teacher	239
guitar	13	professional players or teachers	252

guitar: Each student played all seven pieces in a less reverberant and less noisy environment at home, about one meter away from the recording device (Roland R-07).

3.2 Textual Feedback for Each Performance Recording

As online lessons have become the norm in music colleges due to COVID-19, a similar lesson plan was adopted in our method.

Each teacher wrote one textual feedback for each performance recording, as if you were giving a daily lesson, and each teacher in total wrote 6 (for piano), 20 (for oboe), and 20 or 19 (for guitar) pieces of textual feedback. A total number of textual feedback for each instrument is shown in Table 2. The performances were selected in a counterbalanced manner with the following constraints: each teacher reviewed two performances for each piece, and each student was reviewed by all the teachers throughout the performances. Audio files of performance recordings were sent to each teacher, along with the following introduction: "Please provide textual feedback for each recording assuming the daily lessons." They listened to each recording and either wrote or typed their feedback. For oboe, one piece of textual feedback was lost during the collection process.

3.3 Questionnaire Survey of the Usefulness of the Textual Feedback

We conducted an online questionnaire survey via a crowdsourcing platform[10] to evaluate the usefulness of the textual feedback. We recruited 400 (100 for piano/ 200 for oboe/ 100 for guitar) people who had musical experience outside of school and asked them to provide their demographic informations and answer the question "Do you think that this feedback is useful for future performances?" using a 11-point Likert scale (10: useful – 0: useless). Each participant responded to 46 (for piano), 50 (for oboe), and

[10] https://www.lancers.jp

Table 3. List of Pieces

Instrument	ID	Composer	Piece
piano	n01	F. Chopin	"Tristesse", Op. 10-3
	n02	F. Chopin	"24 Préludes", Op. 28-7
	n03	J. S. Bach	Invention No. 1 in C major, BWV 772
	n04	J. S. Bach	Invention No. 15 in M minor, BWV 786
	n05	L. v. Beethoven	Sonata No. 8 in A flat major, Op. 13
	n06	L. v. Beethoven	Sonata No. 8 in C minor, Op. 13
	n07	R. Schumann	"Traumerai", Kinderszenen No.7, Op. 15
	n08	W.A. Mozart	Sonata No.32 in A major, KV. 331
	n09	C. Debussy	La Fille aux Cheveux de Lin
	n10	C. Debussy	Rêverie
oboe	n01	L. v. Beethoven	Symphony No. 3 in E flat major 'Eroica', Op. 55
	n02	G. A. Rossini	'La Scala di seta' Overture
	n03	F. Schubert	Symphony No. 8 in B minor D.759 'Unfinished'
	n04	J. Brahms	Violin Concerto in D major, Op. 77
	n05	P. I. Tchaikovsky	Symphony No. 4 in F minor, Op. 36
	n06	P. I. Tchaikovsky	"Swan Lake", Ballet Suite, Op.20a
	n07	N. Rimsky-Korsakov	"Scheherazade", Symphonic Suite, Op. 35
	n08	R. Strauss	"Don Juan", Symphonic Poem, Op. 20
	n09	M. Ravel	Le Tombeau de Couperin I.Prelude
	n10	S. Prokofiev	"Peter and the Wolf", Symphonic Tale, Op. 67
guitar	n01	F. Sor	Etude No. 1, Op. 31-1
	n02	F. Sor	Etude No. 5, Op. 35-22
	n03	M. Carcassi	Etude, Op. 60-3
	n04	Anonymous	Romance: Jeux interdits
	n05	F. Tárrega	Lágrima
	n06	L. Walker	Kleine Romanze
	n07	J. S. Sagreras	Maria Luisa

100 (for guitar) randomly selected pieces of textual feedback. Different participants were recruited for each musical instrument.

Hereinafter, in this paper, the average value for each textual feedback will be referred to as its usefulness.

3.4 Annotation of Types of Sentences in Textual Feedback

The purpose of this study was to identify which instructional contents that are significantly more useful for performance students as the more they are mentioned in the textual feedback. An annotation was assigned to each sentence of textual feedback to categorize them by content. Then, we obtained the number of sentences of each content type in each piece of textual feedback. Periods, exclamation marks, or question marks were considered as sentence breaks.

Table 4. Types of Instruction Contents

Types	Definition	Example of sentence
Giving Subjective Information (GSI)	Teacher providing general and/or specific conceptual information based on teacher's subjectivity.	*The tone is soft and comfortable to listen to.*
Giving Objective Information (GOI)	Teacher providing general and/or specific conceptual information based on objectively referable events or concepts.	*Too much arpeggio on the chords in bar 32 would sound unnatural.*
Asking Question (AQ)	Enquiring.	*Is there a problem with the tuning of the instrument?*
Giving Feedback (GF)	Teacher evaluation of a student's applied and/or conceptual knowledge.	*The detailed phrasing of the melody is well expressed.*
Giving Practice (GP)	Providing suggestions of ways to practice a particular passage or discussing a practicing schedule.	*The first step in practicing is to play only the melody.*
Giving Advice (GA)	Giving a specific opinion or recommendation without demonstration or modelling to guide the student's action towards the achievement of certain specific musical aims.	*I think it would be better to be more aware of the larger phrases and not stop the music so much within these phrases.*

The content types were adapted from the works of Simone [37], Carlin [4], and Zhukov [47] as shown in Table 4.[11] Each sentence was annotated as one of these six types of content. If a sentence was judged to consist of descriptions that could be classified as more than one type of content, the sentence was separated using commas. Two annotators annotated all the textual feedback. If their annotations for a sentence differed, they discussed the sentence and settled on a final annotation. The Cohen's Kappa coefficient, which was a statistic to measure inter-rater reliability, was 0.96 for the oboe dataset.

4 Contents that Contribute to the Usefulness of Textual Feedbacks

In this section, we used the usefulness of each textual feedback (Section 3.3) and the number of sentences that meant each content in each textual feedback (Section 3.4) to identify content that significantly improves the usefulness of textual feedback.

[11] Types of "Demonstrating", "Modelling", and "Listening/Observing" were omitted because these actions might be not observed in textual feedback.

4.1 Usefulness

First, this subsection showed demographic data on usefulness of each textual feedback. The average usefulness is following[12]; piano: 6.75 ($\pm$ 2.01), oboe: 7.27 ($\pm$ 2.02)[13], and guitar: 7.15 ($\pm$ 1.90).

The textual feedback with the highest usefulness and the lowest usefulness for the piano are presented below.

The highest rated textual feedback for piano (p06-s02-c21, usefulness: 8.12 $\pm$ 1.68)
You have read the score carefully. The tempo was a little slow compared to the Allegro, so do your best to play faster as you continue to practice.
Check the Es in the third beat of the left hand in the 10th bar because you played it incorrectly.
You played the two-hand staccato in the 17th bar too long, so cut it a little shorter (same in the 78th bar).
You played the left-hand note in the 19th bar by extending it to the first beat of the 20th bar, rather than the whole beat, so you should cut it off properly on beat 1 (same for the left-hand in the 23rd bar).
I don't feel the sforzando in the 33rd and 34th bars at all. Play it with more force (don't hit the keyboard).
Don't extend the left-hand note in bar 40 a whole beat.
The beginning of the fourth beat of the right hand in the 43rd bar was slow.
I understand that you want to rit. from the descending flow of the right-hand note in the 42nd bar, but you should play it without slowing down the tempo as the score shows.
The half note in the 46th bar was long. Since it is staccato, let's play it as long as a quarter note.
I am concerned about the right hand note that comes in without a pause after the fermata in the 61st bar. As you can see when you actually sing it, you always need to breathe to start a new song after such a long note. Be sure to breathe at the eighth rest.

The lowest rated textual feedback for piano (p08-s02-c20, usefulness: 3.75 $\pm$ 2.27)
The melody sounded good and the harmony was well-balanced.
It was a very nice performance.

4.2 Hierarchy of Usefulness

This subsection explores whether the usefulness of text feedback depended on the teacher, the student, or the piece, independent of the musical instruments.

[12] Since the crowdworkers who participated in the questionnaire survey of usefulness were different for each musical instrument, an absolute value comparison of usefulness among musical instruments is not very meaningful.
[13] For oboe, this result was reprint of [29].

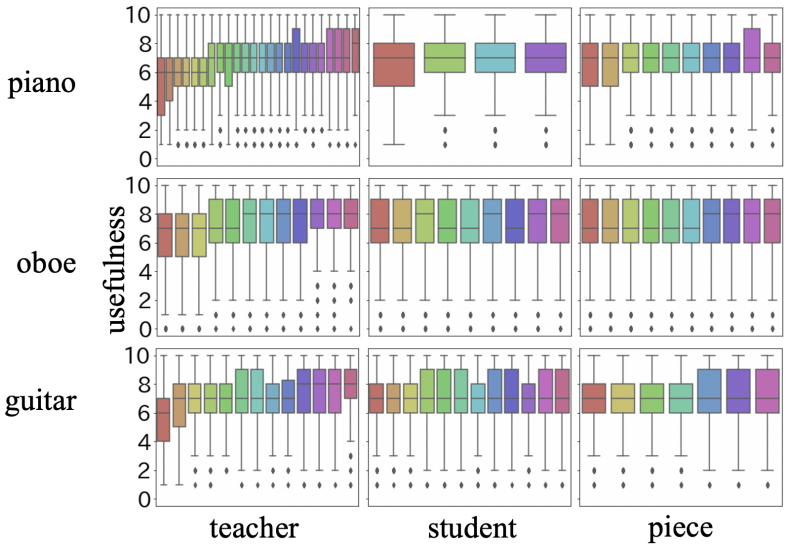

Fig. 1. The average usefulness for each teacher, student, and piece (sorted by usefulness score)

Figure 1[14] shows the average usefulness for each teacher, student, and piece. This result implies that the usefulness of the text feedback differed more by the teacher than by the piece or student.

For teachers, players, and pieces, the intraclass correlation coefficient (ICC) and the design effect (DE)[15] were calculated for each musical instrument.

piano: For teachers, ICC was 0.43 and DE was 3.14. For students, ICC was 0.006 and DE was 1.21. For pieces, ICC was 0.0, and DE was 1.0.

oboe: For teachers, ICC was 0.45 and DE was 9.43. For students or pieces, ICCs were 0.0, and DEs were 1.0[16].

guitar: For teachers, ICC was 0.55 and DE was 11.2. For students or pieces, ICCs were 0.0, and DEs were 1.0.

In summary, independent of musical instruments, the usefulness showed hierarchy among teachers.

4.3 Factors Contributing to the Usefulness of Textual Feedback

For each of the three musical instruments, the content that significantly improves the usefulness of textual feedback was analyzed.

[14] For oboe, the figure was reprint of [29].

[15] DE is a criterion that takes into account both the average number of data in the group and ICC. $DE = 1+(k^*-1)ICC$, where k^* is the average number of data in the group. An ICC greater than 0.05 or a DE greater than two suggested that the data were hierarchical.

[16] For oboe, this result was reprint of [29].

Method Multilevel modeling was conducted to quantitatively analyze the effect of number of sentences annotated as each type of content for each of the three musical instruments. Multilevel modeling enables analysis assuming that the behavior of individual data changes depending on the hierarchy of data. In other words, in this study, not only the change in usefulness among textual feedback but also the influence of the teachers could be analyzed. R 4.1.0 and brms 2.15.0 were used for this multilevel modeling.

In the i-th feedback of the j-th participants, the usefulness of the k-th content U_{ij} is designated as follows:

$$U_{ij} = \alpha + \sum_{k=1}^{6} \beta_k n_{ik} + \eta_k^{(z_{ijkg})} + \sum_{k=1}^{6} \gamma_k^{(z_{ijkg})} n_{ik} + e_{ij}$$

Let α be intercept, k be each type of contents, $\beta_k (k = 1, \ldots, 6)$ be the coefficient of n_{ik}, n_{ik} be the number of descriptions for the k-th category. Here, z_{ijkg} indicates each teacher who wrote the i-th feedback for each musical instrument. g indicates each teacher; $g \in \{1, \ldots, 24\}$ for piano, $g \in \{1, \ldots, 12\}$ for oboe, and $g \in \{1, \ldots, 13\}$ for guitar. $\eta_k^{(z_{ijkg})}$ is the random effect of the teacher on the intercept for the k-th content category of the i-th feedback. $\gamma_k^{(z_{ijkg})}$ is the random effect of the teacher on the coefficient for n_{ik}.

The model parameters were fitted with four Markov chain Monte Carlo chains with 2,000 iterations and 1,000 burn-in samples with a thinning parameter of one. Specifically, we used $\beta_k \sim N(0, 100)$, $\alpha \sim StudentT(3, 0, 2.5)$, and $\sigma_e \sim StudentT(3, 0, 2.5)$ as the prior distributions of the fixed effects, $StudentT(3, 0, 2.5)$ as the prior distribution of SD of random effects, and $LKJCholesky(1)$ as the prior distribution of the correlation matrix between $\eta_k^{(g)}$ and $\gamma_k^{(g)}$ for $k \in \{1, \ldots, 6\}$.

Results Correlation coefficients between the six variables were checked for each musical instrument and all were less than 0.8, so all variables were used in the analysis. For each instrument, the estimates and 95% credible intervals of each content are shown in Table 5. R-hats for all features were under 1.05.

Table 5. Overview of the Results

	piano	oboe	guitar
β_{GSI}	-0.11[-0.24,-0.01]	0.07[-0.09, 0.23]	-0.02[-0.09, 0.05]
β_{GOI}	**0.07 [0.00, 0.14]**	**0.13 [0.06, 0.21]**	**0.12 [0.07, 0.19]**
β_{AQ}	0.05[-0.23, 0.32]	0.41[-1.11, 1.86]	-0.05[-0.34, 0.28]
β_{GF}	0.03[-0.09, 0.15]	**0.14 [0.04, 0.25]**	0.07[-0.01, 0.15]
β_{GP}	-0.08[-0.82, 0.14]	**0.27 [0.10, 0.45]**	**0.13 [0.04, 0.22]**
β_{GA}	**0.16 [0.09, 0.24]**	**0.15 [0.08, 0.22]**	0.08[-0.00, 0.17]

These results showed that the number of sentences that conveyed *GOI* were significantly positive for all of the three instruments. Therefore, the number of sentences

conveying *GOI* significantly contributes to the usefulness of the textual feedback independent of musical instruments. Moreover, the number of sentences conveying *GP* or *GA* significantly contributes to the usefulness of the textual feedback for the two musical instruments.

5 Discussion

In this study, textual feedback on musical recordings of oboe, piano, and guitar pieces were collected and analyzed. We quantitatively found that the usefulness of the textual feedback differed most significantly by teachers independent of musical instruments. Moreover, the number of sentences conveying *GOI* was found to significantly contribute usefulness of the textual feedback independent of musical instruments. In this study, different levels of students and teachers were involved in the collected recordings and textual feedback for each musical instrument (Table 1). Therefore, the instrument-independent results suggested that the results may be independent of the level of the player and the instructor.

Our result that the number of sentences conveying *GOI* significantly contributes to the usefulness of the textual feedback has high generalizability because the results can apply to face-to-face lessons.

In this study, the experiments were conducted only in Japanese. In the future, it will be necessary to conduct comparisons across multiple languages and discuss the differences between languages and cultures [3]. Another limitation was that this study used textual data as verbal information. We cannot deny the possibility that the verbal information in face-to-face speech shows different characteristics from the verbal information in textual data. An exploration of whether feedback should be psychologically supportive or what words should be used should be undertaken in the future.

6 Conclusion

We published the dataset for investigating the use of verbal feedback for three musical instruments. This dataset clarified that the content of text feedback were different between the teachers, and the feedback conveying *giving objective information* was critical for students independent of the musical instrument. In the future, we would like to utilize these findings in the development of educational programs.

References

1. Bayley, J.G., Waldron, J.: “it’s never too late”: Adult students and music learning in one online and offline convergent community music school. Int. J. Music. Educ. **38**(1), 36–51 (2020)
2. Biasutti, M., Antonini Philippe, R., Schiavio, A.: Assessing teachers’ perspectives on giving music lessons remotely during the covid-19 lockdown period. Musicae Scientiae **26**(3), 585–603 (2022)
3. Campbell, P.S.: Lessons from the world: A cross-cultural guide to music teaching and learning. MacMillan Publishing Company (1991)

4. Carlin, K.D.: Piano pedagogue perception of teaching effectiveness in the preadolescent elementary level applied piano lesson as a function of teacher behavior. Ph.D. thesis, Indiana University (1997)
5. Cavitt, M.E.: A descriptive analysis of error correction in instrumental music rehearsals. J. Res. Music. Educ. **51**(3), 218–230 (2003)
6. Chen, Y.A., Yang, Y.H., Wang, J.C., Chen, H.: The amg1608 dataset for music emotion recognition. In: ICASSP. pp. 693–697 (2015)
7. Choi, E., Chung, Y., Lee, S., Jeon, J., Kwon, T., Nam, J.: Ym2413-mdb: A multi-instrumental fm video game music dataset with emotion annotations. In: Ismir 2022 Hybrid Conference (2022)
8. Dickey, M.R.: A comparison of verbal instruction and nonverbal teacher-student modeling in instrumental ensembles. J. Res. Music. Educ. **39**(2), 132–142 (1991)
9. Dorman, P.E.: A review of research on observational systems in the analysis of music teaching. Bull. Counc. Res. Music. Educ. pp. 35–44 (1978)
10. Duke, R.A.: Measures of instructional effectiveness in music research. Bull. Counc. Res. Music. Educ. pp. 1–48 (1999)
11. Duke, R.A., Simmons, A.L.: The nature of expertise: Narrative descriptions of 19 common elements observed in the lessons of three renowned artist-teachers. Bull. Counc. Res. Music. Educ. pp. 7–19 (2006)
12. Foscarin, F., McLeod, A., Rigaux, P., Jacquemard, F., Sakai, M.: ASAP: a dataset of aligned scores and performances for piano transcription. In: ISMIR. pp. 534–541 (2020)
13. Froehlich, H.: Measurement dependability in the systematic observation of music instruction: A review, some questions, and possibilities for a (new?) approach. Psychomusicology **14**(1-2), 182 (1995)
14. Gillman, D., Goyat, U., Kutlay, A.: Teach yourself georgian folk songs dataset: An annotated corpus of traditional vocal polyphony. In: Ismir 2022 Hybrid Conference (2022)
15. Goolsby, T.W.: Verbal instruction in instrumental rehearsals: A comparison of three career levels and preservice teachers. J. Res. Music. Educ. **45**(1), 21–40 (1997)
16. Goto, M., Hashiguchi, H., Nishimura, T., Oka, R.: Rwc music database: Popular, classical and jazz music databases. In: ISMIR. pp. 287–288 (2002)
17. Goto, M., Hashiguchi, H., Nishimura, T., Oka, R.: Rwc music database: Music genre database and musical instrument sound database. In: ISMIR. pp. 229–230 (2003)
18. Hamanaka, M., Hirata, K., Tojo, S.: Gttm database and manual time-span tree generation tool. In: SMC. pp. 462–467 (2018)
19. Hash, P.M.: Remote learning in school bands during the covid-19 shutdown. J. Res. Music. Educ. **68**(4), 381–397 (2021)
20. Hashida, M., Matsui, T., Katayose, H.: A new music database describing deviation information of performance expressions. In: ISMIR. pp. 489–494 (2008)
21. Hashida, M., Nakamura, E., Katayose, H.: Constructing pedb 2nd edition: a music performance database with phrase information. In: SMC. pp. 359–364 (2017)
22. Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.Z.A., Dieleman, S., Elsen, E., Engel, J., Eck, D.: Enabling factorized piano music modeling and generation with the MAESTRO dataset. In: ICLR (2019)
23. Hentschel, J., Neuwirth, M., Rohrmeier, M.: The annotated mozart sonatas: Score, harmony, and cadence. Transactions of the International Society for Music Information Retrieval **4**(1) (2021)
24. Lehmann, A.C., Sloboda, J.A., Woody, R.H., Woody, R.H., et al.: Psychology for musicians: Understanding and acquiring the skills. Oxford University Press (2007)
25. Li, B., Liu, X., Dinesh, K., Duan, Z., Sharma, G.: Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. IEEE Tran. Multimedia **21**(2), 522–535 (2018)

26. Matsubara, M.: Crocus: Dataset of musical performance critique (Jun 2021). https://doi.org/10.5281/zenodo.4748243
27. Matsubara, M., Kagawa, R., Hirano, T., Tsuji, I.: Analysis of the usefulness of critique documents on musical performance: Toward a better instructional document format. In: Ke, H.R., Lee, C.S., Sugiyama, K. (eds.) Towards Open and Trustworthy Digital Societies. pp. 344–353. Springer International Publishing (2021)
28. Matsubara, M., Kagawa, R., Hirano, T., Tsuji, I.: Crocus: Dataset of musical performance critiques: Relationship between critique content and its utility. In: CMMR (2021)
29. Matsubara, M., Kagawa, R., Hirano, T., Tsuji, I.: Useful feedback in asynchronous lessons of music performance: A pilot study of oboes. In: The Journal of the Society for Art and Science. pp. 241–255 (2022)
30. Miragaia, R., Reis, G., de Vega, F.F., Chávez, F.: Multi pitch estimation of piano music using cartesian genetic programming with spectral harmonic mask. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI). pp. 1800–1807. IEEE (2020)
31. Nakamura, E., Saito, Y., Yoshii, K.: Statistical learning and estimation of piano fingering. Information Sciences **517**, 68–85 (2020)
32. Salamon, J.: What's broken in music informatics research? three uncomfortable statements. In: 36th International Conference on Machine Learning (ICML), Workshop on Machine Learning for Music Discovery. Long Beach, CA, USA (2019)
33. Sapp, C.S.: Comparative analysis of multiple musical performances. In: ISMIR. pp. 497–500 (2007)
34. Schedl, M.: The lfm-1b dataset for music retrieval and recommendation. In: ICMR. pp. 103–110 (2016)
35. Silla Jr, C.N., Koerich, A.L., Kaestner, C.A.: The latin music database. In: ISMIR. pp. 451–456 (2008)
36. Simones, L., Schroeder, F., Rodger, M.: Categorizations of physical gesture in piano teaching: A preliminary enquiry. Psychology of Music **43**(1), 103–121 (2015)
37. Simones, L.L., Rodger, M., Schroeder, F.: Communicating musical knowledge through gesture: Piano teachers' gestural behaviours across different levels of student proficiency. Psychology of Music **43**(5), 723–735 (2015)
38. Sturm, B.L.: An analysis of the gtzan music genre dataset. In: ACM Workshop MIRUM. pp. 7–12. MIRUM '12 (2012)
39. Ueda, K., Takegawa, Y., Hirata, K.: Evaluation of a piano learning support system focusing on visualization of keying information and annotation. In: E-Learn. pp. 1198–1204 (2015)
40. Vigliensoni, G., Fujinaga, I.: The music listening histories dataset. In: ISMIR. pp. 96–102 (2017)
41. Visentin, P., Shan, G., Wasiak, E.B.: Informing music teaching and learning using movement analysis technology. Int. J. Music. Educ. **26**(1), 73–87 (2008)
42. Wang, Z., Chen, K., Jiang, J., Zhang, Y., Xu, M., Dai, S., Bin, G., Xia, G.: Pop909: A pop-song dataset for music arrangement generation. In: ISMIR (2020)
43. Weiß, C., Zalkow, F., Arifi-Müller, V., Müller, M., Koops, H.V., Volk, A., Grohganz, H.G.: Schubert winterreise dataset: A multimodal scenario for music analysis. J. Comp. Cult. Herit. **14**(2), 1–18 (2021)
44. Whitaker, J.A.: High school band students' and directors' perceptions of verbal and nonverbal teaching behaviors. J. Res. Music. Educ. **59**(3), 290–309 (2011)
45. Yalcinalp, S., Avci, Ü.: Creativity and emerging digital educational technologies: A systematic review. Turkish Online Journal of Educational Technology-TOJET **18**(3), 25–45 (2019)
46. Zhang, K., Zhang, H., Li, S., Yang, C., Sun, L.: The pmemo dataset for music emotion recognition. In: ICMR. pp. 135–142 (2018)
47. Zhukov, K.: Teaching styles and student behaviour in instrumental music lessons in Australian conservatoriums. Ph.D. thesis, University of New South Wales (2005)