NUFluteDB: Flute Sound Dataset with Appropriate and Inappropriate Blowing Styles

Sai Oshita[1] and Tetsuro Kitahara[1]*

[1]Graduate School of Integrated Basic Sciences, Nihon University
Setagaya-ku, Tokyo, Japan
{ohshita,kitahara}@kthrlab.jp

Abstract. This paper describes a dataset of flute sounds with appropriate and inappropriate blowing styles. The flute is known as a difficult instrument to learn. We, therefore, have been developing a support system that automatically identifies the appropriateness of blowing in flute performances. To develop such a system, we need a dataset that consists of various sounds with various blowing styles, including both appropriate and inappropriate ways, but there are no such datasets. In this paper, we present the dataset that we have been developing. This dataset consists of sounds played by various players with various blowing styles, and also it has annotations of each sound's subjective evaluation.

Keywords: Flute, Dataset, Subjective evaluation

1 Introduction

The flute is an instrument whose sound changes greatly when the breath's direction and the mouth's size are changed. Therefore, it is necessary to carefully control the size of the mouth and the direction and strength of the breath to play the flute with appropriate tone quality. However, although many books on the market instruct how to play the flute, only a few clearly describe these points. Therefore, even if one reads a detailed book, it is not easy for a beginner to listen to their sound and judge how to improve it by themselves.

To facilitate beginners' practice of the flute, we have been developing a system that analyzes users' flute sounds and feeds back on how inappropriate their sounds are and/or why they are inappropriate. To achieve such technologies, we need a lot of flute sounds played by various players with different skill levels.

Several studies have been conducted on flute performance support systems. Yoonchang[1] created a system to judge whether the player was playing appropriate sounds by evaluating the head-tube relationship, air pressure, and fingering from the sounds. Kuroda et al.[2] created a dataset that includes sounds played by a robot and human

* This work was supported by JSPS Kakenhi Nos. JP22H03711 and JP21H03572.

players to control the blow's strength and direction strictly. After creating their dataset called *Good Sounds*, Romani et al.[3] created a system to analyze the acoustic characteristics of flute sounds.

Datasets of flute sounds have also been developed recently. Brum[4] created a dataset commprising performances of four pieces with different directions. Cantos[5] created a dataset of flute sounds to research automatic music transcription; it contains monophonic and polyphonic flute sounds, their MIDI transcriptions, and objective evaluations of the transcription accuracy. Goto et al.[6] created a dataset of sounds of various instrument by different performers with different intensities. In addition, multiple datasets of sounds other than flute performances have been developed [7–10]. However, a dataset has not been developed that includes various sounds played with both appropriate and inappropriate blowing styles with annotations of their subjective evaluation.

In this paper, we describe a dataset we have created for a flute performance support system. This dataset is a combination of flute-playing sounds and their ratings.

2 Dataset

Because this dataset aims to develop a support system of flute practice, the dataset has to include various sounds ranging from novice-level to advanced ones. In addition, each sound should have an annotation representing how it sounds well. Therefore, we can summarize the issues in designing the dataset as follows:

- **The skill level of players**
  Various players ranging from novices to experts should participate in our recording. In particular, asking novice players to participate is essential because such players may hesitate to record their flute sounds, even though it is crucial to analyze sounds played by such players.
- **Playing styles**
  The dataset should include sounds played in inappropriate styles, such as too large mouth, too small mouth, too upward breath, and too downward breath. In particular, it would be adequate to ask skillful players to play in such styles intentionally.
- **The number of sounds to be collected**
  The dataset should include as many sounds as possible. It is helpful to collect sounds on the Web because participants record their sounds without restrictions on the place and time. It was also significant because our lifestyle was strongly influenced by COVID-19 when we made the dataset.
- **Annotations of subjective evaluation**
  To evaluate the appropriateness of flute sounds played by various players on a computer, we need an annotation of the subjective evaluation of each sound. We have to ask sufficiently advanced players to do subjective evaluations to keep the evaluations reliable.

This strategy has some limitations. One is a lack of uniformity in the recording quality. Because participants record their sounds and send them to us via the Web, they are assumed to be recorded via their own devices (such as smartphones). Also, the

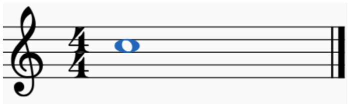
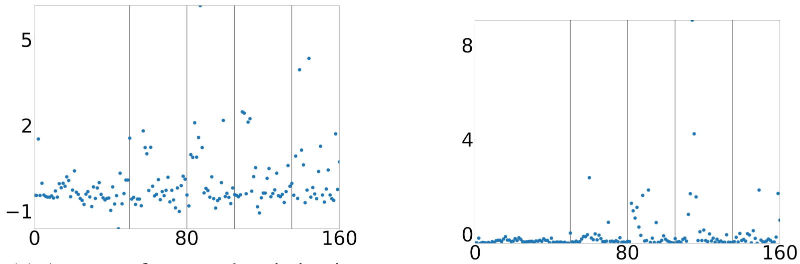


Fig. 1. Note performed by participants



(a) Average of temporal variation in amplitude (d_v)

(b) Average of temporal variation in F0 (d_f)

Fig. 2. Acoustic features of the collected flute sounds. Horizontal: sound ID (from left to right: [Normal], [Large mouth], [Small mouth], [Breath upward], [Breath downward]), Vertical: standardized feature values

recording environment (e.g., the distance between the microphone to the flute) cannot be unified. The lack of uniformity may have negative influences on sound analysis.

Another is that we cannot check if participants follow our instructions. Even if they are asked to play in the ”too large mouth” style, no one can check if they are genuinely opening their large mouth.

Even though it has such limitations, we made a dataset based on this strategy. Below, we mention the details of the dataset. The dataset is available at the following URL:
`https://github.com/5418010saiohshita/dataset`

2.1 Audio recordings

We collected flute sounds played in various blowing styles, including both appropriate and inappropriate ones. As inappropriate ways, we focused on mouth size and breath direction. Due to COVID-19, we asked performers to record their performances themselves and collected them on a crowdsourcing site. The performers were asked to play the score shown in Figure 1 without vibrato. To reduce the burden on individual performers, we asked either of the following two patterns:

1 [Normal] [Large mouth] [Small mouth] [Breath upward]
2 [Normal] [Breath downward]

The details of the performers and the number of collected sounds are listed in Tables 1 and 2, respectively. To compensate for the fact that the sound volume varies depending on the recording conditions, we corrected the amplitudes so that the temporal mean values of the amplitudes are equal.

Figure 3 shows acoustic features of the collected flute sounds: the averages of temporal variations in amplitude and fundamental frequency (F0). Regarding both the amplitude and F0, sounds played in non-normal blowing styles tend to have more considerrable temporal variations.

Table 1. Details of musical experience etc. of performers

Per-formers	Age	Exp.* [yrs.]	Gap in exp. [yrs.]	Non-flute experience	Max non-flute exp. [yrs.]	Self-determined flute level
P01	46	2	5	Trumpet	9	Beginner
P02	34	10	3	piano, Guitar	5	Intermediate near beginner
P03	55	2	0	Piano	7	Beginner
P04	29	3	0	Piano	5	Intermediate near beginner
P05	33	2	1	Tenor sax, soprano sax, alto sax	12	Beginner
P06	51	0.8	2	Piano, alto sax, clarinet, bass clarinet	11	Almost no experience
P07	21	3	10			Intermediate near beginner
P08	21	0.1				Intermediate near beginner
P09	46	5	10	Piano	12	Beginner
P10	16	8				Intermediate near advanced
P11	24	10	2			Intermediate near beginner
P12	30	10	5	Piano	16	Intermediate near advanced
P13						
P14						

Exp.: experience
Empty cells mean unanswered.
*Some performers may have answered years of experience excluding the gap.

Fig. 3. Distribution of subjective evaluation 1 (overall quality) of collected sound. Horizontal: sound ID (from left to right: [Normal], [Large mouth], [Small mouth], [Breath upward], [Breath downward]), Vertical: ratings

2.2 Subjective evaluation

To each sound collected above, we annotated its blowing appropriateness. To obtain such annotations, we conducted subjective evaluations of the collected sounds using a web-based crowdsourcing service. Participants were limited to current or former students of flute majors in music colleges or high school music departments and those who have played the flute for at least 12 months. As a result, six participants listed in Table 3 participated. The number of participations is different among the participants because we allowed them to participate several times as they would like.

When each participant opened the designated web page, 20 randomly selected sounds were displayed. They listened to them individually and entered their answers to the questions in Table 4. The choices for choice-type questions are listed in Tables 5 and 6.

Figure 3 shows the distribution of subjective evaluation 1 (overall quality) for the collected sounds. In general, sounds played in the normal-blowing style were given higher ratings than those in the non-normal-blowing style.

Table 2. Number of Experiments for flute sound collection

Performers	Blowing styles Normal	Larger	Smaller	Upward	Downward	Total
P01	1	1	1	1	0	4
P02	2	1	1	1	1	6
P03	10	5	5	5	5	30
P04	1	0	0	0	1	2
P05	1	1	1	1	0	4
P06	10	5	5	5	5	30
P07	2	1	1	1	1	6
P08	7	4	4	4	4	23
P09	7	7	7	7	6	34
P10	3	3	3	3	3	15
P11	2	1	1	1	1	6
P12	1	1	1	1	0	4
P13	1	0	0	0	1	2
P14	1	0	0	0	1	2
Total	49	30	30	30	29	168

Table 3. Participants (evaluators) for subjective sound evaluation

Participant	# of participation	Flute experience [yrs.]	Non-flute experience
S01	1	3	Sax, piano
S02	10	5	Piano, harp
S03	1	6	Electric organ, percussion, etc.
S04	1	22	Piano, percussion, piccolo, etc.
S05	1	28	Piano
S06	1	28	Piano

Table 4. Questions used in the subjective evaluation

1	Overall quality	Response type 1
2	Clearness of the tone	Response type 1
3	Stability of the intensity	Response type 1
4	Stability of the pitch	Response type 1
5	Smallness of the breathy noise	Response type 1
6	Which in the blowing problems apply? (one or more)	Response type 2
7	Write anything else you noticed	Description

Table 5. Response type 1 for subjective evaluation

1	Below beginner level. Seen as just starting level.
2	Beginner level. There are some areas that need improvement.
3	Intermediate level. Some improvement is needed. In general, the student's performance is satisfactory.
4	Intermediate to advanced level. There are some points to be improved, but the performance is acceptable for an amateur concert.
5	Advanced level. There is nothing to be improved at all.

Table 6. Response type 2 for subjective evaluation

1	Breathing too strong
2	Breathing too weak
3	Mouth size too large
4	Mouth too small
5	Breath too upward
6	Breath too downward
7	No problem
8	I don't know

Table 7. Acoustic features extracted from flute sounds

Feature	Feature description
d_v	Average of time variation of amplitude
d_f	Average of time variation of fundamental frequency
r_v	Amplitude range
r_f	Fundamental frequency range
o_s	Number of harmonic components (including fundamental frequency components) at the beginning of blowing
f_s	Percentage of overtones (non-fundamental components) in all harmonics at the beginning of blowing
n_s	Percentage of overtones in the entire spectrum at the beginning of blowing
o_c	Number of overtones (calculated from the middle interval)
f_c	Percentage of non-fundamental frequency components in all overtones (calculated from the middle interval)
n_c	Percentage of overtone components in the whole spectrum (calculated from the middle interval)

3 Examples of the use of this dataset

In this section, we present examples using the dataset we created [1].

3.1 Predicting subjective evaluation from acoustic features

We conducted the prediction of subjective evaluation from acoustic features. This would help develop support system for flute practice. Here, we used linear regression. From each audio signal included in the dataset, 10 acoustic features listed in Table 7 are extracted. Then, these features are applied to linear regression. In linear regression, the objective variable is subjective evaluation 1 (overall quality), while the explanatory variables are those features. Half data were assigned to the training data and the rest to the test data.

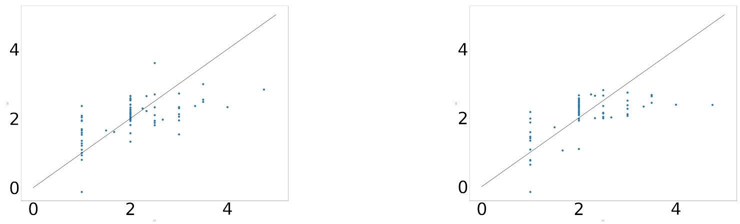
Figure 4 compares the subjective evaluation's predicted and actual values. The figure shows that even though the actual value of the highest subjective evaluation is 4.75, and its predicted value is 2.38. When the outliers are removed, the sounds where the actual subjective evaluation is greater than 3 have lower predicted values than the actual evaluation. The root mean square error (RMSE) of the prediction is 0.670. When the outliers are removed, the RMSE is 0.642.

We also attempted the same prediction with decision trees (DTs) after the subjective evaluation was discretized into two or three classes (that is, we conducted it as two-class or three-class classification). The objective feature and explanatory features are the same as above. Table 8 lists the classification accuracy and the depth of the trees acquired. An example of the trees is shown in Figure 5.

3.2 Predicting blowing styles from acoustic features

When the performers recorded a sound, they were asked the blowing style from [Normal], [Large mouth], [Small mouth], [Breath upward], and [Breath downward]. We

[1] These have been presented in our previous paper [11].



(a) Outliers included (b) Outliers removed

Fig. 4. Actual subjective evaluation (horizontal) and its prediction (vertical) with linear regression

Table 8. Accuracy of predicting subjective evaluation with DT (in parentheses: outliers removed)

Classification	Maximum(Depth 2)	Maximum
Two-class (Lower than 2 / 2 or higher)	0.93 (0.94)	0.93 (0.94)
Three-class (Lower than 2 / 2 to 3 / 3 or higher)	0.83 (0.84)	0.86 (0.84)

attempted the prediction of this blowing style from the acoustic features. We conducted different classification tasks with DTs: two-class [Normal / Other], three-class [Normal / Mouth-size-related / Breath-direction-related], and five-class: each style. Table 9 lists the classification accuracy. An example of the acquired trees is shown in Figure 6.

4 Conclusion

We presented a flute sound database consisting of sounds played in appropriate and inappropriate blowing styles. This dataset is intended to be used for developing a support system of flute practice by analyzing how inappropriate the user's sounds are and why. To help such analysis, we annotated the subjective evaluation to each sound.

In addition, we presented examples of flute sound analysis using our dataset. Even though the prediction of subjective evaluation using linear regression and DTs showed promising results to some extent, the accuracy for predicting blowing styles was low. One possible reason could be that the performer could not strictly control the mouth size and breath direction.

In the future, we would like to improve how to collect sounds. For example, we will ask advanced players to control their mouth size and breath direction strictly and will check them via video recordings. Through this, we would like to develop technologies that help novice flute players improve their skills.

References

1. Yoonchang, H. and Kyogu, L.: Hierarchical Approach to Detect Common Mistakes of Beginner Flute Players, ISMIR, 2014
2. Jin, K.and Gou, K.: Sensing Control Parameters of Flute from Microphone Sound Based on Machine Learning from Robotic Performer, Sensors, 22, 2022
3. Romani, O., Parra, H., Dabiri, D., Tokuda, H., Hariya, W., Oishi, K., and Serra, X.: A Realtime System for Measuring Sound Goodness in Instrumental Sounds, 138th Audio Engineering Society Convention, 2015

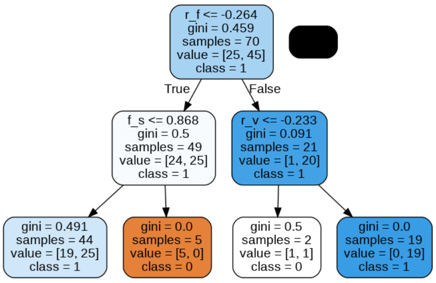


Fig. 5. DT for predicting subjective evaluation

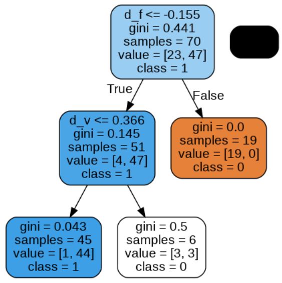


Fig. 6. DT for predicting blowing styles

Table 9. Accuracy of predicting blowing styles with DT (in parentheses: outlier removal)

Classification	Maximum (Depth 2)	Maximum
Two-class (Normal, Others)	0.70 (0.41)	0.71 (0.75)
Three-class (Normal, Oral, Breath)	0.39 (0.31)	0.49 (0.44)
Five-class (Each blowing style)	0.36 (0.19)	0.36 (0.19)

4. Brum, J. P. B.: Traditional Flute Dataset for Score Alignment, 2018, `https://www.kaggle.com/jbraga/traditional-flute-dataset`
5. Elena A. C.: Flute Audio Labelled Database for Automatic Music Transcription, 2018, `https://doi.org/10.5281/zenodo.1408985`
6. Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R. : RWC Music Database: Music Genre Database and Musical Instrument Sound Database, ISMIR, 2003.
7. Ito, K. and Johnson, L.: The LJ Speech Dataset, 2017, `https://keithito.com/LJ-Speech-Dataset/`
8. Vassil, P., Guoguo, C., Daniel, P., and Sanjeev, K.: LibriSpeech: an ASR Corpus Based on Public Domain Audio Books, ICASSP, 2015
9. John, K. and Alan, W. B.: CMU ARCTIC Databases for Speech Synthesis, CMU-LTI-03-177, 2003
10. Paul, B.: Automatic Prosodic Analysis for Computer Aided Pronunciation Teaching, PhD thesis, University of Edinburgh, UK, 1994
11. Oshita, S. and Kitahara, T.: Automatic Classification of Blowing Properness in Flute Sounds, ICA, ABS-0467, 2022