STRUCTURE-INDEPENDENT FACTORS IN EXPRESSIVE TIMING: A PRELIMINARY STUDY ON VIOLIN SOLO PERFORMANCE

Tzu-Yun Hung
National Taiwan Normal University
hong.megan@gmail.com

Yo-Wei Hsiao, Li Su
Institute of Information Sciance, Academia Sinica
willyhsiao,lisu@iis.sinica.edu.tw

ABSTRACT

We investigate the static behavior of individual notes in real-world music performance. Different from the previous works which mostly focused on the dynamical behavior of expressive timing in performance, we assume that expressive timing has a global trend which is independent from musical structure, cadence, and other time-varying factors. An expression-wise study on a violin solo dataset with 11 different music expressions verifies our assumption.

1. INTRODUCTION

In music notation, the relative duration of a note, or the so-called *note value*, is often described in the base-2 numeral system, e.g., we divide a whole note into two half notes, four quarter notes, eight eighth notes, and so on. Such a binarized hierarchical time scale is, however, deformed with musicians' interpretation of music expressions in real-world music performance. Analyzing the tempo or micro-timing in music performance is therefore a central topic in music performance analysis [1–3]. Previous research in this direction has mostly focused on studying how the *time-varying* behavior of tempo and inter-onset interval (IOI) related to different music structures and expressions, performers, and eras [4–6]. On the contrary, relatively few research considered finding the *static*, structure-independent behavior of each individual note in music performance, probably because this approach is less musical.

In this work, we consider one static analysis which investigates *on average* how musicians scale the duration and IOI of a note given its note value and expression. Specifically, we assume that performers' interpretation of musical time not only is time-varying, but also incorporates structure- and time-independent factors. We are then interested in the research questions: 1) how musicians' interpretation of note value deviates from the base-2 time scale in general? and 2) how such deviation varies with different music expressions? Our research methods and results are reported as follows.

2. METHOD

Our study is based on four recordings selected from the SCREAM-MAC-EMT dataset, which is a dataset for music expression research and contains violin solo recordings performed by 40 musicians with 11 different music expression terms in 10 pieces [7]. To reduce the variation of composition style, the four pieces are selected works of Antonio Vivaldi's *The Four Seasons*: two of them are the first and second movement from *Spring*, and the other two are the first and third movement from *Autumn*.

The process we obtained the onset and offset time of each note in the four selected pieces is described as follows. First, we use Tony [8], a software for monophonic note transcription, to retrieve the note onsets and offsets. The transcription results were checked and revised by human by adding missed notes and removing false alarm notes. However, the human-edited results is not free from the inevitable discrepancy in subjective perception. We further adopt a calibration process which moves each onset event to its nearest peak of the spectral flux feature [9, 10].

Secondly, the offset events are calibrated by the energy of the signal between each two consecutive onsets. The energy curve is estimated for each frame. The offset event which lies in between two consecutive onsets is the time which makes the ratio of the average energy before the offset and the average energy after the offset be largest. In other words, we assume that an offset is the time that marks the greatest drop of energy within an inter-onset interval.

The note value depends on the beat value indicated by the denominator of the meter. To simplify our discussion, we convert note value into beats, for example, a 1-beat note can be a quarter note in the meter 4/4, or a half note in 2/2. Based on the given musical notation in the score, we split the available notes into note duration classes defined by four note values (k = 0.25, 0.5, 1, 2 beats) over the 11 expressions, dotted notes are excluded due to the insufficient sample sizes. Based on this, we compute the *average beat length* (i_1) and *average note duration per beat* (d_1) for each recording. This is done by weighted average over the IOI (for i_1) and duration (for d_1) of each note, where the weight of a 1-beat note is 1, the weight of a 0.5-beat note is 2, and the weight of a k-beat note is $1/k$. As a result, i_1 and d_1 are the *expected* beat length and note duration, respectively, of a 1-beat note in the recording. One can therefore expect that, for example, the expected IOI of a 2-beat note is $2i_1$.

To figure out how the timing in the performance deviates from this binary time scale, we consider the following three normalized temporal features for each note: 1) normalized k-beat note duration (d_k), which is the average duration of all the k-beat notes in the recording divided by kd_1 of the same recording; 2) normalized k-beat IOI (i_k), which is the average IOI (ignore rest) of all the k-beat notes in the recording divided by ki_1 of the same recording, and 3) normalized k-beat note coverage (c_k), which is the average duration of all the k-beat notes in the recording divided by ki_1 of the same recording. Then, we averaged the values of all the 40 performance recordings and the results are denoted as $\bar{d}_k$, $\bar{i}_k$ and $\bar{c}_k$, respectively. A value higher than 1 means that the note terms are longer (or slower) than the expected length; otherwise, it indicates that the length in the real performance is shorter (faster) than the average.

3. RESULTS

Figure 1 illustrates the result of normalized k-beat note duration ($\bar{d}_k$, left), note IOI ($\bar{i}_k$, middle), and note coverage ($\bar{c}_k$, right). From the left of Figure 1, we observe that in general, $\bar{d}_k$ increases with k except for the 0.25-beat notes; even $\bar{d}_2 > 1$ for all expressions. In other words, the attack-decay-sustain-release (ADSR) envelope is further lengthened for long notes. Another notable factor is that in these pieces, the 0.5-beat notes are usually followed by 0.25-beat notes. To make brief notes heard clearly, performers tend to emphasize the 0.25-beat notes. From the middle of Figure 1, we observe a clear trend that the normalized k-beat IOI, $\bar{i}_k$, decreases with the nominal note length k. This global trend demonstrates a perceptual scale of music time in which a $2k$-beat note tends to be performed shorter than two k-beat notes. This is possibly because performers tend to emphasize short notes while being less patient in filling the time intervals of long notes. Then, the right of Figure 1 shows that the normalized note coverage ($\bar{c}_k$) goes long for long notes; this implies that musicians incline to fill in the long notes and make short notes shorter in duration. In summary, musicians tend to accelerate long notes while filling in their duration (i.e. *legato*), and slow down short notes while leaving some rest time (i.e. *staccato*).

It should be noted that these trends also depend on the limitation of human's motor system as well as the difficulty of the music piece. The 3rd movement of *Autumn* contributes greatly to the lengthened $\bar{d}_{0.25}$ and $\bar{i}_{0.25}$, and this might also be resulted from its difficulty. With its tempo in *Presto* and with hasty switching between strings, violinists are less capable of performing the piece with a regular tempo. Besides, without bowing and dynamic indication on the score, violinists would choose their own way to perform. Some of them changed the bowing in every note, which would make the performance much harder.

By comparing different expressions, first, we observe that the normalized k-beat values ($\bar{d}_k$, $\bar{i}_k$ as well as $\bar{c}_k$) of *Maestoso* , *Tranquillo* and *Scherzando* tend to deviate from 1, while in contrast, the values of *Cantabile*, *Grazioso* and *Agitato* tend to be close to 1. Most importantly, we found that *non-expressive* expression deviates

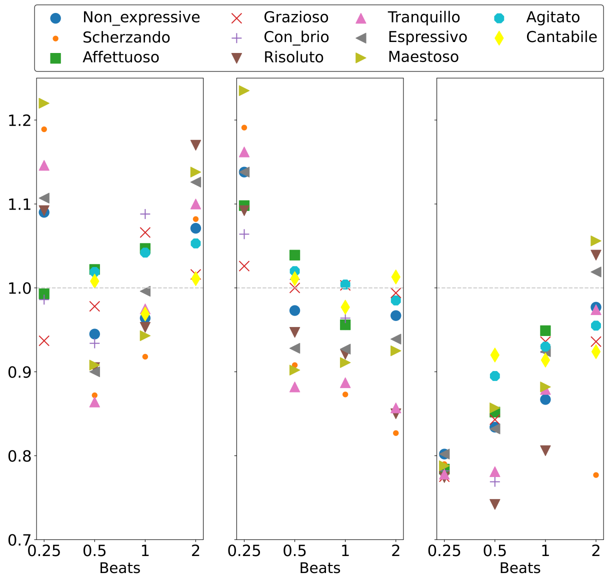

Figure 1: The normalized lengths versus the nominal beat length (k) for various musical expression terms. Left: normalized k-beat note duration. Middle: normalized k-beat IOI. Right: normalized k-beat note coverage.

more than the three expressions. Another notable observation is that the normalized 0.25-beat duration and IOI of *Maestoso* are the longest among all. This phenomenon does reveal the interpretation of *Maestoso*: as being "dignified" and "stately,"[1] musicians tend to perform pieces with more deeply (i.e. strong energy) and expressively (i.e. slowly) emotion, and in the way short notes have to be longer than expected. Another interesting thing is that *Scherzando*, an expression quite different from *Maestoso*, also exhibits long $\bar{d}_{0.25}$ but the shortest $\bar{d}_2$; this implies that accelerating long notes and slowing down short notes is a characteristic of *Scherzando*. In some expressions such as *Affettuoso* and *Agitato*, the normalized k-beat note duration increases with k. In *Agitato*, $d_k > 1$ for all k, which means that the notes with k not in our discussion (e.g., dotted notes) in *Agitato* are commonly shorter than average. Other expression terms mostly display an irregular trend with k.

4. CONCLUDING REMARKS

In this study, we have compared three types of temporary independent, note-level timing features in performance among 11 different music expression terms. Observations on the differences among these expression terms have been made, including the special characteristics of the *non-expressive* performance, which actually demonstrates more expressive timing than some other expression terms. Besides, different from the structural- and temporal-dependent analysis approaches in previous research, here we show that a static analysis provides novel perspectives, and also sheds light on the research of expressive timing in music performance.

[1] https://dictionary.cambridge.org/dictionary/italian-english/maestoso

5. REFERENCES

[1] A. Lerch, C. Arthur, A. Pati, and S. Gururani, "An interdisciplinary review of music performance analysis," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, pp. 221–245, 2020.

[2] C. E. Cancino-Chacón, M. Grachten, W. Goebl, and G. Widmer, "Computational models of expressive music performance: A comprehensive and critical review," *Frontiers in Digital Humanities*, vol. 5, p. 25, 2018.

[3] P. Desain and H. Honing, "Does expressive timing in music performance scale proportionally with tempo?" *Psychological Research*, vol. 56, no. 4, pp. 285–292, 1994.

[4] D.-J. Povel, "Temporal structure of performed music: Some preliminary observations," *Acta Psychologica*, vol. 41, no. 4, pp. 309–320, 1977.

[5] M. Rector, "Historical trends in expressive timing strategies: Chopin's etude, op. 25 no. 1," *Empirical Musicology Review*, vol. 15, no. 3-4, pp. 176–201, 2021.

[6] H. Honing, *Structure and Interpretation of Rhythm in Music*. Academic Press, 2012.

[7] P.-C. Li, L. Su, Y.-H. Yang, A. W. Su *et al.*, "Analysis of expressive musical terms in violin using score-informed and expression-based audio features." in *ISMIR*, 2015, pp. 809–815.

[8] M. Mauch, C. Cannam, R. Bittner, G. Fazekas, J. Salamon, J. Dai, J. Bello, and S. Dixon, "Computer-aided melody note transcription using the tony software: Accuracy and efficiency," in *Proceedings of the First International Conference on Technologies for Music Notation and Representation*, May 2015, accepted.

[9] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on speech and audio processing*, vol. 13, no. 5, pp. 1035–1047, 2005.

[10] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx). Maynooth, Ireland (Sept 2013)*, vol. 7, 2013.

[11] M. Tian, G. Fazekas, D. A. A. Black, and M. Sandler, "On the use of the tempogram to describe audio content and its application to music structural segmentation," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 419–423.