ANALYSIS AND SYNTHESIS OF THE VIOLIN PLAYING STYLE OF HEIFETZ AND OISTRAKH

Chi-Ching Shih, Pei-Ching Li, Yi-Ju Lin, Yu-Lin Wang, Alvin W. Y. Su*

SCREAM Lab., Department of CSIE,
National Cheng-Kung University
Tainan, Taiwan
P78021015@mail.ncku.edu.tw

Li Su

Music and Culture Technology Lab.,
Institute of Information Science,
Academia Sinica
Taipei, Taiwan
lisu@iis.sinica.edu.tw

Yi-Hsuan Yang

Music and Audio Computing Lab.
Research Center for IT Innovation,
Academia Sinica
Taipei, Taiwan
yang@citi.sinica.edu.tw

ABSTRACT

The same music composition can be performed in different ways, and the differences in performance aspects can strongly change the expression and character of the music. Experienced musicians tend to have their own performance style, which reflects their personality, attitudes and beliefs. In this paper, we present a data-driven analysis of the performance style of two master violinists, *Jascha Heifetz* and *David Fyodorovich Oistrakh* to find out their differences. Specifically, from 26 gramophone recordings of each of these two violinists, we compute features characterizing performance aspects including articulation, energy, and vibrato, and then compare their style in terms of the accents and legato groups of the music. Based on our findings, we propose algorithms to synthesize violin audio solo recordings of these two masters from scores, for music compositions that we either have or have not observed in the analysis stage. To our best knowledge, this study represents the first attempt that computationally analyzes and synthesizes the playing style of master violinists.

1. INTRODUCTION

Music performance is composed of two essential parts: the compositions of the composers (scores) and the actual interpretation and performance of the performers (sounds) [1]. A piece of music can be interpreted differently depending on the style of the performers, by means such as extending the duration of accents or manipulating the energy of individual notes in a note sequence, etc.

Performance aspects that characterize the style of a violinist may include interpretational factors such as articulation and energy, and playing techniques such as shifting, vibrato, pizzicato and bowing. For example, for the articulation aspect of music, the time duration of individual note (DR) and the overlap in time between successive notes (or key overlap time, KOT) in the audio performance have been found important in music analysis and synthesis [2]. These two features characterize the speed of the music and the link between successive notes in a note sequence. Maestre *et al.* [3] used them to develop a system for generating an expressive audio from a real saxophone signal, whereas Bresin *et al.* [4] used them to analyze the articulation strategies of pianists.

Research has also been done concerning the techniques of violin playing. For instance, Matsutani [5] analyzed the relationship between performed tones and the force of bow holding. Ho *et al.* [6] proposed that we can achieve good intonation while performing vibrato by maintaining temporal coincidence between the intensity peak and the target frequency. Baader *et al.* [7] presented a quantitative analysis of the coordination between bowing and fingering in violin playing. These studies all collected scientific data, such as the location of fingerprint on bow, the finger shifting data, pitch values, and velocity of finger, that may help amateur violinists improve their performing skills.

The analysis of musician style is also used in sound synthesis to improve the quality of synthetic music pieces. For example, Mantaras *et al.* [8] studied compositional, improvisational, and performance AI systems to characterize the expressiveness of human-generated music. Kirke *et al.* [9] studied factors including testing status, expressive representation, polyphonic ability, and performance creativity, to highlight the possible future directions for media synthesis. Li *et al.* [10] analyzed how expressive musical terms of violin are realized in audio performances and proposed a number of features to imitate human-generated music.

The style differences among violinists have been studied for a long time [11, 12, 13, 14]. For instance, Jung [15] presented an analysis of the playing style of three violinists, *Jascha Heifetz*, *David Fyodorovich Oistrakh* and *Joseph Szigeti*. He argued that the precision in the performance of Heifetz gives listeners the feeling that Heifetz is "cold" and "unemotional." In contrast, Oistrakh was described as "warm" and "capable of communicating emotional feelings." To some extent, the style of Heifetz and Oistrakh can be considered as two extremes of the spectrum, making the comparison between them interesting.

In this paper, we present a step forward into expressive performance analysis and synthesis of violin music by learning from the music of these two masters, Heifetz and Oistrakh. To this end, we first compile a new dataset consisting of some manual annotations of 26 excerpts from famous violin concertos that these two masters have both played (cf. Section 2). As perceptually Heifetz and Oistrakh are quite different in terms of the *velocity* and *accent* of their music, we choose to focus on the analysis of the articulation, energy, and vibrato aspects of the music in this dataset and propose a few features (cf. Section 3.1). In particular, we find that DR and KOT are indeed useful in characterizing the differences in the accents of their music (cf. Section 3.2). Then, we combine the vibrato synthesis procedure proposed by Yang *et al.* [16] and the proposed features to obtain synthetic sounds that imitate the style of these two violinists (cf. Section 4). We hope that this endeavor can make us closer to the dream of reproducing the work of the two masters via expressive synthesis.

2. DATASET

To compare the style of Heifetz and Oistrakh, we choose the music pieces that both of them have played before. In particular, we focus on the work recorded in the prime of their lives, for they are considered to be representative of their style. In addition, we pick only

Table 1: *Dataset information.*

Composer	Music Name	Movement	Bars
L. V. Beethoven	*Violin concerto in D major, Op. 61*	I	96, 97-98, 100-101, 128-130, 130-131, 134-136, 137-139
		II	77, 81-82, 83
J. Brahms	*Violin concerto in D major, Op. 77*	I	115-118, 124-127, 129-132
P. I. Tchaikovsky	*Violin concerto in D major, Op. 35*	I	28-29, 31-33, 39-41, 105-106
M. C. F. Bruch	*Scottish Fantasy, Op.46*	II	97-101, 107-109
E. V. A. Lalo	*Symphonie espagnole in D minor, Op. 21*	I	41-51, 66-69
J. Sibelius	*Violin concerto, Op. 47*	I	26-31
		III	124-127, 128-130, 133-135, 146-152

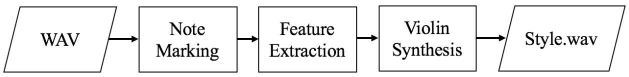

Figure 1: *Flowchart of playing style analysis and synthesis system.*

the violin solo parts to keep the influence from the background accompaniment minimal. Table 1 provides information about the resulting 26 excerpts we select from six violin concertos. We note that, because we have the performance of both the two masters for each of the excerpts, we have in total 52 audio recordings.

The original sources of these music pieces are gramophone recordings.[1] For analog to digital conversion, we use the Technics SL-1200 turntable, the phono amplifier ASR Basis phono preamplifier, and the Audio-Technica AT440MLa Dual Moving Magnet Cartridge audio interface. The resulting wave files are recorded by Adobe Audition, with 32-bit depth and 48 kHz sampling rate. The average length of these excerpts is around 6 seconds.

To investigate the performance difference between Heifetz and Oistrakh, we annotate the excerpts by hand with detailed information about the onset, offset, legato, accent, bar, note type and tempo. The process of annotation takes over four months to complete. This dataset represents to our best knowledge the first manually annotated dataset on the masterpieces of Heifetz and Oistrakh.

Annotating the accent and note type gives us more information on the note level. An accent is an emphasis on a particular note using louder sound or longer duration. As people may have different ways emphasizing notes, characterizing the accents are important. Except for the accent marks on scores, people may also emphasize a non-accent in their performance. Therefore, the criterion of regarding a note as an accent is not based on the audio data, but is according to the wedge-shaped markings on the score, the stressed beats such as one and three in common time (quadruple), and our judgment while listening to the excerpts.

The note type, on the other hand, helps us differentiate notes of different lengths in our analysis of articulation and energy. For simplicity, we consider the following six types of notes: whole note, half note, quarter note, eighth note, sixteenth note, and thirty-second note. For any note of other types, we assign its note type to be the one with the next longest duration. For instance, a dotted eighth note will be considered as a quarter note in our analysis.

[1] For information regarding label and released year of the gramophone recordings, see `http://screamlab-ncku-2008.blogspot.tw/2017/03/synthesis-result-of-analysis-and.html`. Although we use this dataset to synthesize music, it can also be used for different purposes, such as for education. For example, students can understand what kinds of interpretations these two masters use by listening and analyzing the same music pieces. Besides, students can try to play the violin while listening to one specific excerpt, to feel the diverse tempo between the two masters and to imitate their distinct playing style.

3. METHOD

Figure 1 shows the diagram of the proposed system, whose goal is to extract style-related features and use them for expressive synthesis. Given the audio waveform of an excerpt converted from a gramophone recording, we first mark and segment the notes of the excerpt by hand, for automatic methods for score-to-audio alignment may suffer from the influence of the white noise of gramophone recordings, and the interference of background accompaniments such as cello or the second violin. After getting the onset and offset information from the manual annotations, we calculate note-level features such as duration, energy contour and vibrato extent of every individual note (cf. Section 3.1). Then, according to the scores, we compare the features extracted from the performances of Heifetz and Oistrakh for the same excerpt, with respect to musical notations such as accent, legato (slur) and dynamics (cf. Section 3.2). As a result, we have two sets of statistical parameters that respectively characterize the style of Heifetz and Oistrakh. Depending on which set of statistical parameters we choose in the final synthesis stage, we hope to synthesize audio performance in the style of either Heifetz or Oistrakh (cf. Section 4).

3.1. Feature Extraction

We consider three types of note-level features: articulation, energy, and vibrato. These features are computed for each note. These note-level features are aggregated (pooled) into excerpt-level ones by taking the average across 12 types of note groups: accents and non-accents for each of the six types of notes (i.e. 2×6). We also refer to the excerpt-level features as the statistical parameters.

3.1.1. Articulation

Articulation features, such as DR and KOT, are related to the music tempo and the performers' expressive intentions. Therefore, they are important in playing style analysis [2]. For example, a sequence of short-duration notes without silent intervals between them might make audiences stressful or excited; a sequence of lively music with fixed and short intervals gives the impression of being delightful and lovely, and a long-duration note sequence creates a feeling of peace.

As illustrated in Figure 2, DR represents the duration of a note and we calculate the DR for the n-th note by:

$$DR_n = Offset_n - Onset_n \, , \tag{1}$$

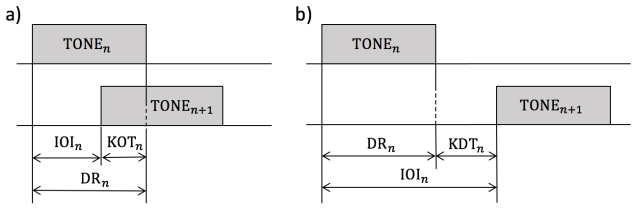

Figure 2: *Illustration of duration (DR), key overlap time (KOT), key detached time (KDT), and inter onset interval (IOI) for (a) overlapping and (b) non-overlapping successive notes $TONE_n$ and $TONE_{n+1}$. For (b), we set $KOT_n = -KDT_n$.*

where $Offset_n$ and $Onset_n$ are the offset and onset times of this note. The KOT, on the other hand, represents the overlap between two successive notes. Higher KOT indicates faster transition between notes. We calculate KOT for the n-th note by:

$$KOT_n = Offset_n - Onset_{n+1}\,. \tag{2}$$

If there is no overlap in time between the n-th note and the next note, we can calculate the key detached time (KDT):

$$KDT_n = Onset_{n+1} - Offset_n\,. \tag{3}$$

In such cases, we consider KOT as the negative KDT.

The excerpt-level features of DR and KOT for an excerpt are composed of the mean value of DR and KOT for accents and non-accents for each type of notes that presents in the excerpt. For example, the excerpt shown in Figure 3 only contains eighth and sixteenth notes. Therefore, the articulation parameters of this excerpt are computed by averaging the note-level DR and KOT values of the eighth note accents, eighth note non-accents, sixteenth note accents, and sixteenth note non-accents, respectively.

Please note that we do not normalize DR and KOT by the tempo of each excerpt, because we consider the tempo also as an indicator of the expression style of a performer.

3.1.2. Energy

Energy is an essential characteristic to distinguish playing style. Performers might follow the dynamic notations of music score in their interpretation. However, oftentimes they would also choose to emphasize a particular note, phrase, or chord by playing it louder, according to their own opinion, to convey different expressions. In this paper, the note-level energy features are the mean energy of each individual note and the mean energy contour for each of the six types of notes. Besides, an excerpt-level feature, accent energy ratio, is also considered.

Before calculating the energy features, we have to carry out the energy normalization due to the different recording environments and equipment of the recordings. The normalization is performed by dividing the energy of an excerpt by its maximum value, making the range of the values $[0, 1]$. After that, we use the short-time Fourier transform with Hanning window to calculate the energy contour of each note, based on its corresponding fundamental frequency F_0, expressing in dB scale. To be specific, given the spectrogram $M(t, k)$ of a note for each time frame t and frequency bin k, the *energy contour* of the note with respect to F_0 is calculated by:

$$EC_{F_0}(t) = 20\log(M(t, F_0))\,. \tag{4}$$

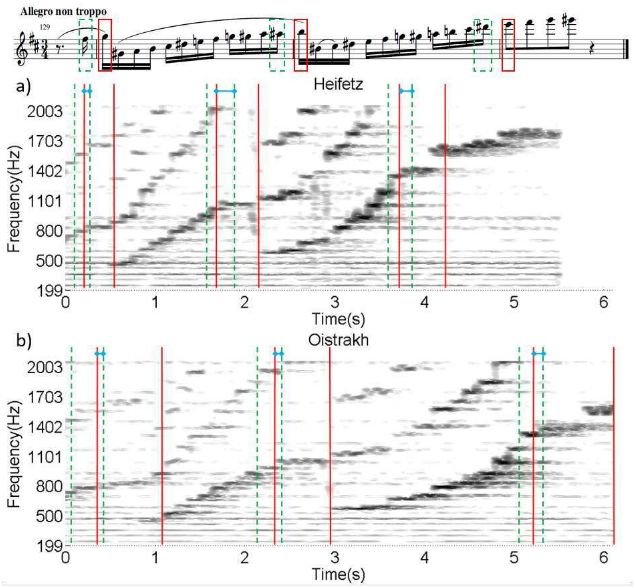

Figure 3: *The spectrograms of Brahms' Violin Concerto, Mov. I, bars 129-132 performed by (a) Heifetz and (b) Oistrakh, with red solid vertical lines and green dashed lines indicating the onset and offset times of accents (red) and their preceding notes, which are typically non-accents (green), respectively. We can see that the KOTs (blue) of (a) Heifetz are larger than those of (b) Oistrakh.*

Then, the energy normalization is applied again for the $EC_{F_0}(t)$ of each note, to keep the maximal peak value equal to one. Lastly, the mean energy contour for a note type is computed by averaging the energy contours of all the notes belonging to that note type.

Moreover, the mean *energy* of an individual note E_n is calculated by summing the amplitudes of the spectrogram across all the frequency bins, expressed in dB scale, divided by the note length:

$$E_n = \frac{20\log(\sum_k M(t,k))}{note\ length}\,. \tag{5}$$

Finally, an excerpt-level feature, *accent energy ratio* (AER), representing the energy ratio of accents to non-accents within an excerpt, is calculated by:

$$AER = \frac{(\sum_{i=1}^{p} E_{A_i})\ /\ p}{(\sum_{j=1}^{q} E_{NA_j})\ /\ q}\,, \tag{6}$$

where E_A is the mean energy of an accent, E_{NA} is the mean energy of a non-accent, p is the number of accents, and q is the number of non-accents. For calculating $EC_{F_0}(t)$ and E_n, we use a frame size of 1,024 samples and a hop size of 64 samples.

3.1.3. Vibrato

Vibrato, the frequency modulation of fundamental frequency, is another essential element in violin performance. We consider the two common note-level features, *vibrato rate* (VR) and *vibrato extent* (VE). VR represents the rate of vibrato, i.e. the number of periods in one second. VE represents the frequency deviation between a peak and its nearby valley. We adopt the vibrato extraction process presented by Li *et al.* [10]. To determine whether a

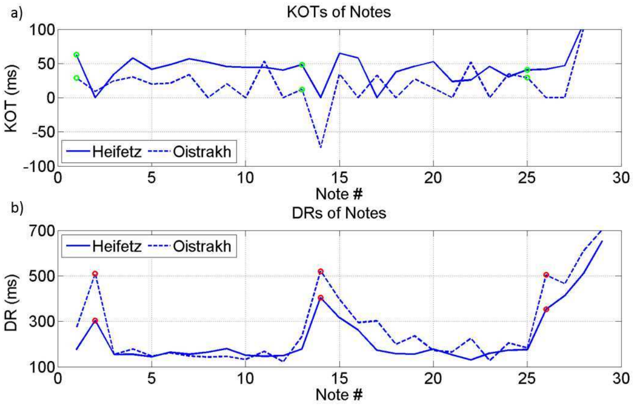

Figure 4: *(a) KOT and (b) DR values of each individual note of the same excerpt used in Figure 3. The circles in (a) indicate the notes right before the accents, while in (b) the circles mark the accents.*

note has vibrato, we require that its VR is between 3 and 10 Hz, VE is between 5 and 100 cents, and note length is over 125ms. The parameters of VR and VE of each excerpt is calculated as the mean value over all the vibrato notes. Moreover, we concern the excerpt-level feature, *vibrato ratio* (vibRatio), also adopted by Li *et al.* [10], to measure the percentage of vibrato notes within an excerpt. The frame size and hop size are the same as those used in computing energy-related features.

3.2. Difference between Heifetz and Oistrakh

The musical notations, accent and legato (slur), are considered in the comparison of articulation and energy features between Heifetz and Oistrakh. Specifically, we focus on the comparison of the KOTs of the notes right before accents, the DRs of accents, and the average energy curve of legato groups. Two excerpts with obvious characteristics are taken as running examples below.

3.2.1. KOTs of the notes right before accents

We observe that the KOTs of the notes right before the accents, which are typically non-accents, are important factors to distinguish Heifetz from Oistrakh. Let's take Brahms' *Violin Concerto*, Mov. I, bars 124-127 as an example. As shown in Figure 3, we mark three accents and the notes right before them by red solid lines and green dashed lines, respectively. According to equation (2), we mark the values of KOTs in blue horizontal lines. It can be seen that the KOTs of Oistrakh are smaller than those of Heifetz. This observation is actually in line with our subjective listening experience: Oistrakh often halts for a moment before playing the accents, while Heifetz, on the contrary, plays without a stop. Figure 4(a) shows the KOT of each note and the circles represent the KOTs of the notes before accents. Apparently, not only such KOT values but also the mean KOT of Heifetz are higher than those of Oistrakh, implying that Heifetz performs in faster tempo.

3.2.2. DRs of accents

From Figure 3, we can also observe for accents, the performance by Oistrakh has larger DRs than that by Heifetz. Besides, according to the legato notation of bar 130 on the score and its corresponding spectrogram, we can see that the DR of the accent takes a large proportion of the legato group, compared with those of the non-accents. In addition, from Figure 4(b), where we use circles to represent the DRs of accents, we have similar finding. This indicates that Oistrakh tends to prolong the duration of accents in comparison to Heifetz. As for the non-accents, both Heifetz and Oistrakh use similar speed to perform them.

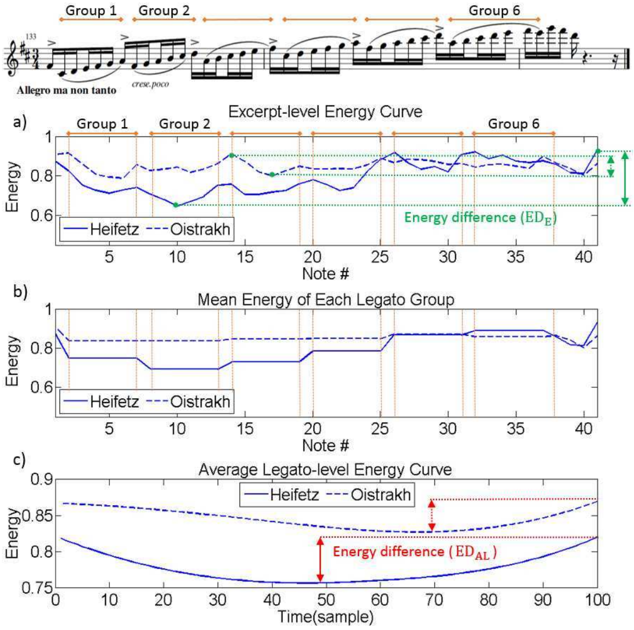

Figure 5: *Energy features for Sibelius' Violin Concerto, Mov. III, bars 133-135: (a) excerpt-level energy curve, (b) mean energy of each legato group, and (c) average legato-level energy curve of the six legato groups.*

3.2.3. Average energy curve of legato groups

In addition to the energy contour of an individual note, the excerpt-level energy curve is also an important factor in playing style analysis. Let's take Sibelius' *Violin Concerto*, Mov. III, bars 133-135 as an example this time. As shown in Figure 5(a), the average energy of each individual note constitutes quite different energy curves for the performances of Heifetz and Oistrakh. We mark the maximal energy differences in green according to the maximal peak value and the minimal valley value within each of these two excerpt-level energy curves. We can see that Heifetz demonstrates an obvious contrast between the energy of the first two bars, whereas Oistrakh uses nearly the same energy for the two bars.

To analyze the energy variation during a set of notes, we use the legato notations on the score and partition all the notes of this excerpt into six legato groups and one non-legato group, which includes the first note and the last three notes. Figure 5(b) shows the mean energy of the six legato groups, from which we also see the gradual tendency of energy increases from the performance of Heifetz, which corresponds to the *crescendo* notation.

It is hard to say anything from the resulting legato-level energy curves, for they seem to differ legato by legato. To simplify our comparison between the style of the two masters, we take the average of the energy curves for all the legato groups and show the resulting average legato-level energy curve in Figure 5(c). After doing so, we can see that the average legato-level energy curve of Heifetz is more convex than that of Oistrakh, implying higher variation of the legato-level energy curves of Heifetz.

Furthermore, we come up with an interesting observation by examining every legato-level energy curve of all the ten excerpts in our dataset that have legato notations on the scores. We find that, when the number of notes of a single legato group is *less than eight*, the legato-level energy curve tends to be a *convex* one. In contrast, when the number of notes of a single legato group is *more than eight*, the legato-level energy curve is *concave*. We will use this finding as a *convex-concave style rule* in our synthesis of the performance of the two masters.

In short, the other two energy-related features that are useful in style-specific expressive synthesis are the maximal *energy difference of an excerpt* ($\mathrm{ED_E}$) and the *energy difference of average legato-level energy curve* ($\mathrm{ED_{AL}}$). The range of the $\mathrm{ED_E}$ and $\mathrm{ED_{AL}}$ values is $[0, 1]$ due to the energy normalization.

4. SYNTHESIS AND RESULTS

Our experiment on expressive synthesis contains two tasks. The first task is to synthesize excerpts that are part of our dataset (i.e. excerpts that have been observed in the analysis stage), in order to test the feasibility of our proposed features and parameters to simulate the fine structure of the original sound. Here, we take Brahms' *Violin Concerto*, Mov. I, bars 129-132, and Sibelius' *Violin Concerto*, Mov. III, bars 133-135, both of which are the two excerpts taken as examples in the previous section. To receive a synthetic sound that looks like the original, the statistical parameters are from the excerpt itself. In other words, we calculate the parameters of each of the two excerpts independently, and apply them to synthesize the playing style of Heifetz or Oistrakh.

The second task is to synthesize an excerpt that is outside of the dataset, Beethoven's *Violin Sonata No.5*, Op.24, Mov. I, bars 1-10, which is also known as *Spring*. The parameters, however, are from the entire dataset, because we do not have any prior knowledge on the expressive parameters of *Spring*. In other words, we want to test via this task whether a computer is able to imitate the two violinists' playing style on an excerpt, which is not collected in the dataset.

The sources are obtained from the part of violin in the RWC database [17], and we particularly select the notes which have no vibrato. Therefore, the synthetic versions can be considered as if the two violinists use another violin to perform the same excerpts.

4.1. Synthesis method

The synthesis process is as follows:

STEP 1 Find the DRs and the KOTs for every accent and every non-accent for each type of notes.

STEP 2 Decide the energy contour of each individual note based on note type.

STEP 3 Decide the mean energy for every accent and every non-accent based on the specific synthetic energy curve of the synthetic target.

STEP 4 Decide which notes are vibrato and which notes are not.

In the first task, the first two steps are straightforward, because the synthetic parameters, the DR, KOT, and energy contour of each individual note, are from the original excerpt itself. In particular, to simplify the procedure, we cut the length of source, according to the specific DR for every accent and non-accent for every note type. As for STEP 3, the fine structure synthetic energy curve is somewhat complicated from the computed parameters. Based on the analysis of excerpt-level and legato-level energy curves, we know the parameters of $\mathrm{ED_E}$ and $\mathrm{ED_{AL}}$ as well as the convex-concave style rule of that excerpt. However, we need to determine the energy difference between successive legato groups and the energy of non-legato groups.

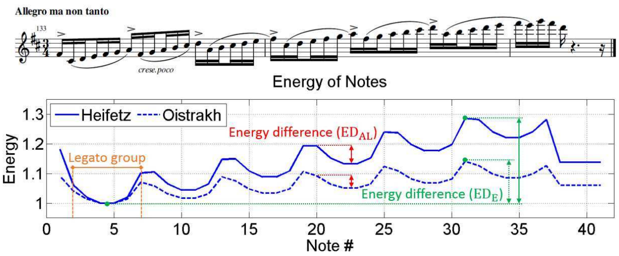

Figure 6: *Synthetic energy curve of Sibelius' Violin Concerto, Mov. III, bars 133-135.*

We have an interesting finding by investigating the excerpts, which are composed of several legato groups with increasing pitch height; that is, both violinists tend to play with increasing energy level among the legato groups for such excerpts. In other words, the mean energy of a legato group increases if its mean pitch is higher than its previous legato group, and vice versa. Therefore, we simply utilize this finding to create an energy curve that is either descending or ascending. Besides, to avoid a sudden change of synthetic energy curve between legato and non-legato groups within an excerpt, we assign the energy of every note of non-legato groups to be the mean value over all the legato groups.

In our implementation, the *synthetic energy curve* of an excerpt is computed as follows:

1. Set the curve as a flat one, implying that the energy of all the notes is the same at the beginning.
2. Decide the tendency of the curve according to the mean pitch of each legato group, and the maximal energy difference of the curve based on the $\mathrm{ED_E}$ value.
3. Decide the trend for every legato group according to the convex-concave style rule, and the energy difference of each legato group based on the $\mathrm{ED_{AL}}$ value.
4. Decide the energy for every note of non-legato groups by taking the average across all the notes of legato groups.
5. Modify the energy for every accent via multiplied by AER value.

Moreover, we set the minimal valley value of each synthetic energy curve to be one, making all the synthetic sounds at the same baseline. Figure 6 shows the synthetic energy curve of Sibelius' *Violin Concerto*, Mov. III, bars 133-135.

Next, in STEP 4, following Yang *et al.* [16], we decide whether a note is a vibrato note based on the vibRatio value. That is, all the notes within an excerpt are sorted in descending order of duration, and the top longest notes are assigned to have vibrato. The exact number of vibrato notes within an excerpt is calculated by the multiplication of the vibRatio value and the total number of

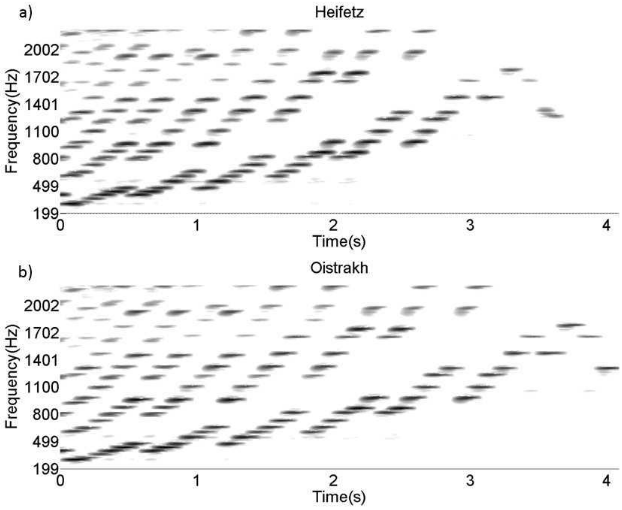

Figure 7: *Synthesis result of Sibelius' Violin Concerto, Mov. III, bars 133-135.*

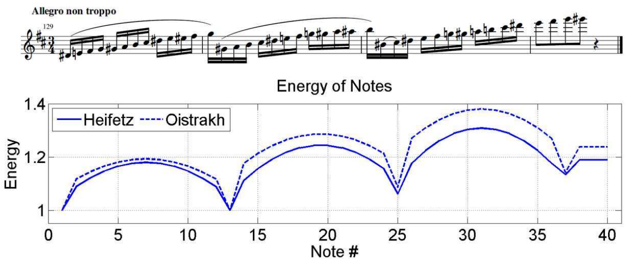

Figure 8: *Synthetic energy curve of Brahms' Violin Concerto, Mov. I, bars 129-132.*

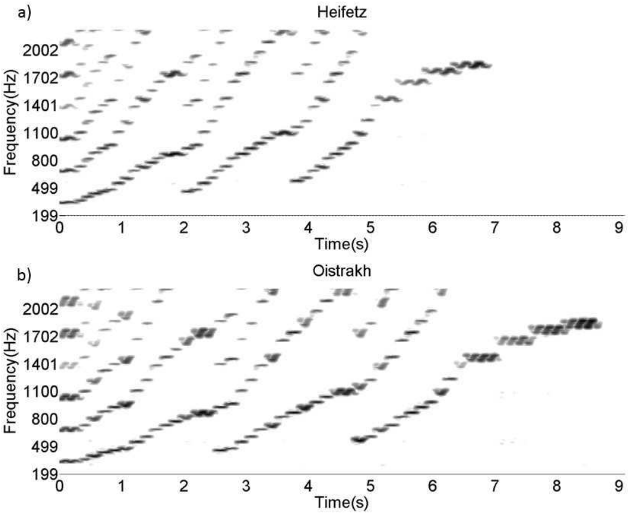

Figure 9: *Synthesis result of Brahms' Violin Concerto, Mov. I, bars 129-132.*

notes in such excerpt. We adopt the vibrato manipulation procedure proposed by Yang *et al.* [16] to synthesize a vibrato note. It is implemented by partitioning a non-vibrato note into a sequence of short fragments, shifting the pitch of each fragment via the phase vocoder to fit the specific vibrato contour, which is constructed by the VR and VE values, and overlapping and adding the fragments to obtain a smooth and natural vibrato sound. Finally, we obtain a synthesis result based on above four steps.

In the second task, there is no prior knowledge on the expressive parameters of *Spring*, which is outside of the dataset, so we take the average of each parameter across all the excerpts at the beginning. However, we find that the synthetic version does not resemble the two performers at all, because the characteristics of some excerpts are diverse, affecting the statistical parameters. To solve this problem, we narrow down the scope of excerpts to those, whose beat per minute (BPM) is around 120, according to the *Allegro* notation on the score of *Spring*. Specifically, we choose 18 excerpts from the dataset, take the mean value of each parameter over such ones, and then apply them to the synthesis system. The synthesis process is the same as the first task.

4.2. Synthesis result

Figure 7 shows the synthesis result of Sibelius' *Violin Concerto*, Mov. III, bars 133-135, whose corresponding synthetic energy curve is shown in Figure 6. The characteristic of accents of both the masters are not obvious; however, Heifetz has clear properties of the KOT, $\mathrm{ED_E}$, and $\mathrm{ED_{AL}}$, implying that Heifetz uses faster tempo and stronger strength to perform this excerpt than Oistrakh.

The score and synthetic energy curve of Brahms' *Violin Concerto*, Mov. I, bars 129-132, are shown in Figure 8. Although the excerpt only has two legato notations on the score, we consider the remainder notes, except for the last four notes, as one legato group, resulting in three legato groups and one non-legato group. There are three concave trends on the synthetic energy curve, because the number of notes of all the three legato groups is more than eight. Besides, Oistrakh has higher energy in his performance, based on the larger $\mathrm{ED_E}$ value. As illustrated in Figure 9, Heifetz performs in faster tempo, especially in the three sequences of short-duration notes, due to his larger KOT value. On the other hand, Oistrakh has strong properties of accents and vibrato notes, especially the last four ones. The former are due to his larger DR and smaller KOT of the notes before accents, and the latter are due to his larger VR, VE, and vibRatio.

Figure 10 shows the score and synthetic energy curve of the *Spring*. To simplify the synthesis procedure, we make an exception to the definition of legato notations of the bar 9; that is, all the notes of this bar are regarded as one legato group. Therefore, except for the first two bars and the last one, the others have convex trends on the synthetic energy curve, according to the convex-concave style rule. Moreover, Heifetz plays louder than Oistrakh based on the larger $\mathrm{ED_E}$ value. The synthesis result of the first four bars is shown in Figure 11, in order to clearly observe the spectrogram. Apparently, Heifetz performs in faster tempo based on the larger KOT. Although the properties of vibrato between the two masters are similar during these four bars, Oistrakh actually has larger vibRatio, indicating that Oistrakh has more vibrato notes within this excerpt than Heifetz does.

In addition, we adopt some post-processings for the original recordings and the synthesis results, to compare them. The frequency band which is under the lowest note of an excerpt, of each original recording is removed, to avoid the influence of the back-

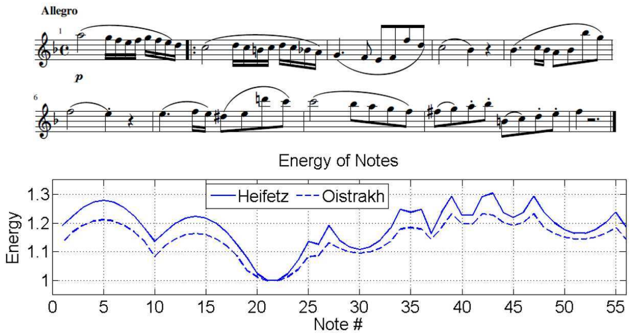

Figure 10: *Synthetic energy curve of Spring.*

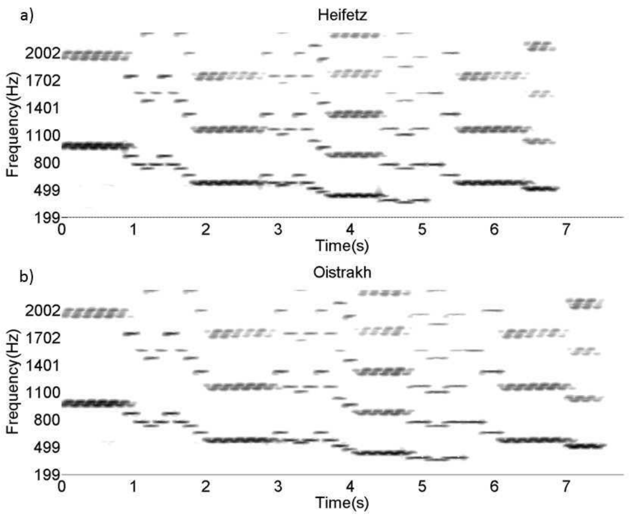

Figure 11: *Synthesis result of Spring. To show the detail of spectrogram, it only contains bars 1-4.*

ground accompaniment. Besides, the energy of each synthesis result is modified to be the same as the original version, whose lower frequency band has also been removed. This way, the audience can focus on the comparison between the original and synthetic sounds in terms of the playing style of the two masters. The sound samples can be found online. [2]

5. DISCUSSION

In this study, we give a quantitative analysis method on the performance of Heifetz and Oistrakh, using the style-related features. Then, we apply these features in expressive synthesis of the playing style of the two performers. As a first attempt to use signal-level features, score information as well as detailed annotations in analyzing this problem, we list some technical limitations of this work which could be left as future work.

The first limitation is the size of the dataset. Since there are only the 26 excerpts from 6 violin concertos in the dataset, it is not enough to cover all kinds of musical style, playing techniques and expressions. Besides, only 10 excerpts can be used in the legato analysis, and some of the energy changes in these excerpts are even constricted because of the symbol on score, which prevents us from observing the rule of energy change. Also, limited size of accent data reduces the reliability of accent modeling between the two violinists, because only the KOT and DR have such statistics. This problem also occurs in the statistics of note type. Furthermore, the dataset only contains 2 excerpts with staccato notes, making it hard to properly model the staccato notes in *Spring*; the parameters have to be calculated from the non-staccato notes.

The second limitation is the capability of vibrato-related features in the synthesis stage. Since we have not fully analyzed the vibrato features of the two masters, the vibrato features are not mentioned in Section 3.2. Although vibrato is an important factor in violin performance, the choice of vibrato notes, the onset time, the vibrato amplitude, etc., have not been considered in this paper. Instead, we compute only the mean vibrato extent and the vibrato rate, and choose the vibrato notes based on the vibrato ratio in the original excerpts. How to better design useful vibrato features and to assign them in the synthesis stage is left as future work.

We point out two further issues in the synthesis step: energy curve construction and accent selection. In the energy analysis, we do not have enough data to classify the energy curves with various shapes. We can only describe the tendency with rules by number of notes and pitch, and synthesize the curve by the difference of maximal and minimal values, instead of the average energy curve models. This way, in turn, causes a stiff energy curve, so that it needs to be improved. For the accent selection, our criterion, which is the first and third beats of a bar (it will only be the first beat when in a triple meter), actually does not yield a realistic synthetic version, especially for *Spring*. The reason is that the energy of the last three bars is not smooth due to the interruption of the energy of accents. Therefore, we would like to find a better selection criterion of accent in the future.

In short, for future work, we will expand the dataset and reconstruct it with no background noise by inviting violinists to imitate Heifetz and Oistrakh, and incorporate more information of energy, vibrato, and accent into expressive synthesis. It is also hoped that the same methodology can be applied to other violinists other than Heifetz and Oistrakh.

6. CONCLUSION

In this study, to our best knowledge, we compile the first manually annotated dataset on the masterpieces of Heifetz and Oistrakh. The annotation contains note-level information such as onset, offset, accent, legato, note type, and so on. We have presented the features distinguishing Heifetz and Oistrakh, and a method to synthesize music pieces based on the analyzed data. The process of expressive synthesis makes use of three signal-level features, including articulation, energy and vibrato, as well as including information from score like accent and legato group. Our result demonstrates the distinct difference between the two performers with respect to the duration, key overlap time, and accent.

7. ACKNOWLEDGMENTS

The authors would like to thank the Ministry of Science and Technology of Taiwan for its financial support of this work, under contract MOST 103-2221-E-006-140-MY3.

[2] `http://screamlab-ncku-2008.blogspot.tw/2017/03/synthesis-result-of-analysis-and.html`

8. REFERENCES

[1] Sandrine Vieillard, Mathieu Roy, and Isabelle Peretz, "Expressiveness in musical emotions," *Psychological research*, vol. 76, no. 5, pp. 641–653, 2012.

[2] Roberto Bresin, "Articulation rules for automatic music performance," in *Proc. of the International Computer Music Conference 2001 (ICMC)*, 2001.

[3] Esteban Maestre, Rafael Ramírez, Stefan Kersten, and Xavier Serra, "Expressive concatenative synthesis by reusing samples from real performance recordings," *Computer Music Journal*, vol. 33, no. 4, pp. 23–42, 2009.

[4] Roberto Bresin and Giovanni Umberto Battel, "Articulation strategies in expressive piano performance analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart's sonata in G Major (K 545)," *Journal of New Music Research*, vol. 29, no. 3, pp. 211–224, 2000.

[5] Akihiro Matsutani, "Study of relationship between performed tones and force of bow holding using silent violin," *Japanese journal of applied physics*, vol. 42, no. 6R, pp. 3711, 2003.

[6] Tracy Kwei-Liang Ho, Huann-shyang Lin, Ching-Kong Chen, and Jih-Long Tsai, "Development of a computer-based visualised quantitative learning system for playing violin vibrato," *British Journal of Educational Technology*, vol. 46, no. 1, pp. 71–81, 2015.

[7] Andreas P Baader, Oleg Kazennikov, and Mario Wiesendanger, "Coordination of bowing and fingering in violin playing," *Cognitive brain research*, vol. 23, no. 2, pp. 436–443, 2005.

[8] Ramon Lopez De Mantaras and Josep Lluis Arcos, "AI and music: From composition to expressive performance," *AI magazine*, vol. 23, no. 3, pp. 43, 2002.

[9] Alexis Kirke and Eduardo Reck Miranda, "A survey of computer systems for expressive music performance," *ACM Computing Surveys (CSUR)*, vol. 42, no. 1, pp. 3, 2009.

[10] Pei-Ching Li, Li Su, Yi-Hsuan Yang, and Alvin WY Su, "Analysis of expressive musical terms in violin using score-informed and expression-based audio features.," in *Proc. of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, 2015, pp. 809–815.

[11] Rafael Ramirez, Esteban Maestre, Alfonso Perez, and Xavier Serra, "Automatic performer identification in celtic violin audio recordings," *Journal of New Music Research*, vol. 40, no. 2, pp. 165–174, 2011.

[12] Eitan Ornoy, "Recording analysis of JS Bach's G Minor Adagio for solo violin (excerpt): A case study," *Journal of Music and Meaning*, vol. 6, pp. 2–47, 2008.

[13] Heejung Lee, *Violin portamento: An analysis of its use by master violinists in selected nineteenth-century concerti*, VDM Publishing, 2009.

[14] Fadi Joseph Bejjani, Lawrence Ferrara, and Lazaros Pavlidis, "A comparative electromyographic and acoustic analysis of violin vibrato in healthy professional violinists," 1990.

[15] Jae Won Noella Jung, *Jascha Heifetz, David Oistrakh, Joseph Szigeti: Their contributions to the violin repertoire of the twentieth century*, Ph.D. thesis, Florida State University, 2007.

[16] Chih-Hong Yang, Pei-Ching Li, Alvin WY Su, Li Su, and Yi-Hsuan Yang, "Automatic violin synthesis using expressive musical term features," in *Proc. of the 19th International Conference on Digital Audio Effects (DAFx)*, 2016.

[17] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, "RWC music database: Database of copyright-cleared musical pieces and instrument sounds for research purposes," *Transactions of Information Processing Society of Japan(IPSJ)*, vol. 45, no. 3, pp. 728–738, 2004.