DIGITAL MUSICOLOGY ASSIGNMENT 2

Joris Monnet[1] **Yutaka Osaki**[1] **Xingyu Pan**[2]
Yiwei Liu[2]
[1] Computer Science, EPFL, Switzerland
[2] Digital Humanitites, EPFL, Switzerland

ABSTRACT

This study aims to compare and analyze the differences between MIDI and human performances and to create an expressive MIDI file based on this analysis. We have taken Chopin's Nocturne Op. 9 No. 2 and compared the original MIDI file with a human pianist Leung's rendition [1], measure by measure.

In the Introduction (Task A), we documented our auditory observations of the midi and performed the type. We noted how human performers convey the emotion and structural intent of the music.

For the modeling part (Task B), we developed a model based on our detailed observations in task A which modify unperformed MIDI files to emulate human expressiveness, with attention to accented notes, dynamic transitions, and rhythm adjustments.

In the discussion section (Task C), we reflected the model in simulating human performance details and proposed potential weakness for model to refine. [1]

1. INTRODUCTION

Three main ways for a human to interpret a piece were discovered: timing, dynamics, and articulation. The musical score [2] of the pieces has also been used to better visualize the raw midi interpretation.

Figure 1: Bars 1-18: timing of left-hand arpeggios

The variation in timing between the notes of the left-hand arpeggios is apparent in measures 1-18. In Fig. 1, we see that the introduction of this work is marked by these arpeggio patterns. The human performance varies in timing, playing the initial and final notes of the arpeggios swiftly, and the intervening notes more slowly, which is different from the uniform play of the MIDI performance.

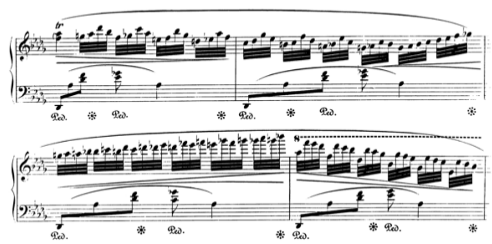

Figure 2: Bars 19-22: crescendos and diminuendos

Figure 3: Bars 51-55: dynamics mark: p and pp

Figure 4: Bars 61: dynamics mark: dim

Dynamics significantly influence the interpretation of measures 19-26, as shown in Fig. 2, and also in measures 35-36, where the score is marked with various dynamic instructions such as crescendos and diminuendos. In a human performance, these dynamics are not only observed but are also often enhanced by pairing crescendos with a slight acceleration, and pairing diminuendos with a gentle deceleration. Additionally, the human-performed version adheres to the pianissimo (pp) and piano (p) markings in the score in Fig.3, and the diminuendo (dim) markings in Fig. 4, guiding the reduction in sound, are all observed. This dynamic shaping, linked with the tempo, enhances the expressive quality of the music, and is absent in the MIDI version, which maintains a consistent volume and pace which is distinctly flat.

[1] https://github.com/JorisMonnet/DM_Assignment2

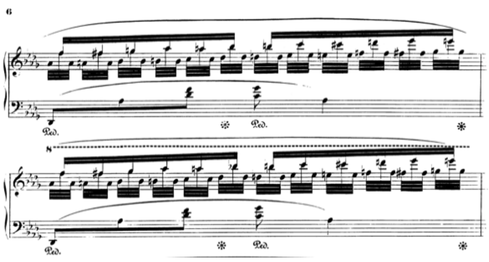

Figure 5: Bars 41-42: Accent

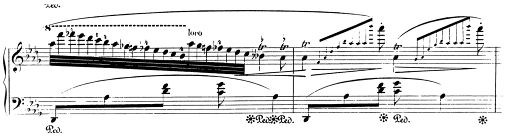

Figure 6: Bars 43-44: Trill

Figure 7: Bars 15-18: Appoggiatura

Figure 8: Bars 53-54: Accent mark

Articulation is notably more expressive in the human-performed version. In measures 40-42 (Fig. 5) and 45-46, where the human player emphasizes the middle, higher-pitched note of each triadic group. Also, the player executes the accent markings "<" (Fig. 8). Furthermore, the extensive use of the sustain pedal, applied twice per measure throughout most of the measures, enhances the resonance and flow of the music. Additionally, the treatment of appoggiaturas indicated in Fig. 7 and trills as shown in Fig.6, adds to the interpretative richness. These aspects which are absent in the MIDI version, shape the phrasing and are particularly noticeable in human performance.

These elements: timing fluctuations of arpeggios, rhythmic emphasis on melodic peaks and articulations, demonstrate the human performer's interpretative approach to translating the score into a emotive musical experience.

2. MODEL

The model simulates the nuances of human performance at different levels (timing, dynamics, and articulation) based on the observation in introduction part.

2.1 Timing

For the observation of articulating arpeggio patterns with varying note durations, function "adjust durations for specific measure" is designed to emulate the nuanced timing variations. It modifies the timing of notes in a measure and applies new note durations to track 1(left hand), and then adjusts track 0'(right hand) note timings to synchronize with these changes, ensuring the duration remains constant.

2.1.1 Acceleration

To capture the tempo acceleration, function "accelerate measure" is designed to adjust the velocity of the MIDI notes—this would control their duration. The function manipulates the score by proportionally adjusting the length of notes within the given measure across specified tracks. The last note in the measure plays at a rate determined by the 'accelerate rate' parameter

2.2 Dynamics

2.2.1 Crescendo and Diminuendo

Functions "change dynamics crescendo measure" and "change dynamics decrescendo measure", are tailored to simulate the effect of crescendo and diminuendos within a measure as observed. They create a crescendo and decrescendo, respectively, within a specified measure in the music stream. They adjust the dynamic marking from a starting level (e.g., "p" for soft) to an ending level (e.g., "f" for forte) across the measure's notes based on the number of distinct note offsets. This simulates the gradual increase or decrease in volume across the measure.

2.2.2 Dynamics curve

The function "classical dynamics shape" crafts a more sophisticated dynamic curve to create a normal crescendo-decrescendo (swell) shape to each measure. This function produce a smooth increase to a midpoint and a symmetric decrease to the end of the measure. The crescendo portion uses a quadratic function to smoothly increase the volume from a minimum to a maximum, while the decrescendo uses a reversed quadratic curve for a smooth decrease back to the minimum volume. This creates a more natural and gradual dynamic swell. Randomness is introduced to each note's velocity to add variability, avoiding the mechanical feel of a strictly uniform dynamic change.

2.3 Articulation

2.3.1 Accent

Function "increase volume of higher notes in track", "increase volume of highest note in triples" and "accentuate highest note in measure" is tailored based on the observation of the accents. They can boosts the volume of the higher-pitched note by a specified amount within specific measures of a given track and handling both individual notes and chords effectively. The first function aims to increase the pitch of the higher note in a pair of adjacent notes; the second function is designed to increase the pitch of the highest note in a group of three contiguous notes; while the third function extends its scope to an entire measure, specifically raising the pitch of the highest note within that measure.

2.3.2 *Trill*

Function "apply trill to hand note" is applied to achieve the trill effect to a selected note. It calculates the necessary number of trill notes based on the given trill speed and duration, and inserts these alternating notes starting from the specified note, thereby simulating the rapid oscillation between two pitches that characterizes a trill.

2.3.3 *Pedal*

For the pedal effect that we observed, the "add pedal event" and "apply pedal to measures" functions are integrated. The former inserts a sustain pedal action into the music at a precise location. It creates a MIDI "ControlChange" event to represent the pedal action that sets it to a value of 127 for pedal down or 0 for pedal up, and decides whether to press or release the pedal based on the 'is pedal down' parameter. The latter adds pedal presses and releases across several measures, suited for a 6/8 rhythm that adds pedal-down events at the start of each beat (at 1/8) and releases at 3/8, then applies it again at 4/8 and releases at 6/8, to simulate the natural pedaling a pianist would use to sustain or mute notes.

2.3.4 *Appoggiaturas*

To simulate appoggiaturas, the function "change duration in measure" is tailored to alter the lengths of specific notes within a given measure, customizing their durations to reflect the expressive timing variations that a human performer would naturally introduce.

3. DISCUSSION

The midi file generated by our model shows improvement in all three areas identified during Task A. However, as anticipated, there are several ways in which it can be further refined.

Regarding articulations, the "add pedal event" and "apply pedal to measures" functions successfully integrate sustain pedal actions into the MIDI performance, using MIDI "ControlChange" events set at 127 for pedal down and 0 for pedal up. However, this binary approach doesn't capture the nuanced gradations of pedal pressure and release that a human performer can achieve. Additionally, our trills come off as somewhat mechanical, but they still show improvement over the raw MIDI file, which lacks these details. We could further enhance trills by shaping the onset of their inner notes. Moreover, we adjusted the duration of ghost notes in appoggiaturas, significantly distinguishing our model from the raw MIDI.

The raw MIDI file's stark absence of dynamic variation was quite surprising. Nevertheless, our model have successfully introduced a range of dynamic changes, significantly enhancing the expressiveness of the piece, as illustrated by the contrasting wave-forms of velocities seen in Fig.9, Fig.10 and 12. We first developed a function to handle crescendos, decrescendos, and accents, yet at times, the result sounded somewhat mechanical. To mitigate this, we incorporated a dynamic contour that mimics a low-high-low pattern within regular measures, thus bringing the sound much closer to that of a human performance. As this function stay quite on the same values, to avoid monotony and maintain a natural feel we even added some randomness. Overall, the implementation has proven to be very effective. In terms of improvement, he Dynamics of human modeled MIDI (Fig.11) displays less variability and the changes in velocities seem more structured, where the ranges of dynamics are more controlled, and there aren't as many peaks and valleys, compared to the human performed (Fig.10) one.

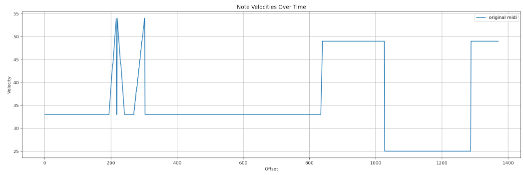

Figure 9: Dynamics of unperformed MIDI

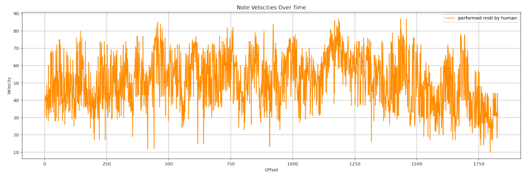

Figure 10: Dynamics of human performance recording

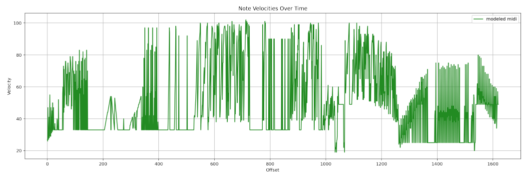

Figure 11: Dynamics of human modeled MIDI

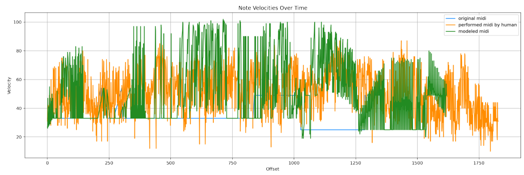

Figure 12: Dynamics comparison

In terms of timing, our model has room for enhancement. We can tell that our improvements are a step up from the raw MIDI, but there's still a gap compared to a real person playing. Slowing down, accelerating and changing speed throughout the piece is quite a remarkable talent and doing it in a pleasant way is far from trivial. We successfully implemented some functions to mimic these handling of timings in the beginning of the piece, but it should be completed for the rest and need a careful listening to work properly. Relevant literature, such as [3], provides insight for these artful executions.

Ultimately, enhancements could be made by adopting a multi-layered approach to humanizing the MIDI file, as suggested in [4], which applies modifications across three tiers: the score, performance, and audio levels. This methodology could further bridge the gap between the digital and the human touch in our MIDI renderings.

4. CONCLUSION

In conclusion, this paper has presented a detailed analysis and computational modeling approach to bridge the gap between MIDI performances and human renditions of music, using Chopin's Nocturne Op. 9 No. 2 as a case study. The comparison between the original MIDI file and a human performance highlighted significant differences in timing, dynamics, and articulation, emphasizing the expressive nuances inherent in human interpretation.

Through the development of a computational model, we aimed to simulate human-like expressiveness in MIDI performances. Our model successfully incorporated adjustments for timing variations, dynamic changes, and articulative nuances observed in human performances. However, there remains room for improvement, particularly in refining timing fluctuations and enhancing the naturalness of dynamic and articulative changes.

Future work could focus on further refining the computational model to capture more subtle nuances of human performance, potentially through the integration of machine learning techniques or more sophisticated algorithms. Additionally, expanding the scope of analysis to include a broader range of musical pieces and performers could provide deeper insights into the complexities of musical expressiveness. Overall, this research contributes to the ongoing efforts in digital musicology to bridge the gap between technology and human creativity in musical expression.

5. AUTHOR CONTRIBUTION

Task A: All
Task B: All
Task C: All
Report: All
Code Cleaning: All
Notebook: All

6. REFERENCES

[1] F. Foscarin, A. McLeod, P. Rigaux, F. Jacquemard, and M. Sakai, "ASAP: a dataset of aligned scores and performances for piano transcription," in *International Society for Music Information Retrieval Conference (ISMIR)*, 2020, pp. 534–541.

[2] "Imslp complete original score," http://conquest.imslp.info/files/imglnks/usimg/6/69/IMSLP851157-PMLP2636-tiff2pdf_(124).pdf.

[3] S. Ji, J. Luo, and X. Yang, "A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions," 2020.

[4] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, "This time with feeling: Learning expressive musical performance," 2018.