DIGITAL MUSICOLOGY ASSIGNMENT 3

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

ABSTRACT

This paper investigates the interpretation and perception of musical phrases within a digital musicology context, focusing on **Moment Musical n°1** by Schubert. Through detailed analysis of MIDI velocity, dynamic contours, and rhythmic patterns, we explore how phrase boundaries are articulated in both symbolic scores and human performances. Our research involves the development of a computational model [1] designed to identify phrase boundaries, which we then test on a larger dataset to evaluate its effectiveness. Understanding phrases is a key component to be able to create human-like rendition of music by computer.

1. INTRODUCTION

In a musical performance, the concept of a 'phrase' can be likened to a sentence in spoken language, marked by its coherent expression of ideas and emotions.

Phrases can be characterize in terms of performance attributes. One of the defining attributes is the contour of its dynamics. In the piece we chose, the "velocity" in MIDI has significant impact on the perception of musical phrases. Phrases often follow a crescendo and decrescendo pattern, indicating a rise and fall in loudness, respectively. This dynamic shape can convey emotional intensity and is closely linked to tempo, where faster passages might be louder, and slower passages softer, mimicking physical movements like breathing or walking. [1] As shown in Fig1, the dynamics follows this kind of contour. Research [2] also demonstrated variations in loudness—whether increasing or decreasing—profoundly affect how listeners perceive and differentiate these musical 'sentences'.

[1] https://github.com/JorisMonnet/DM_Assignment3/tree/main

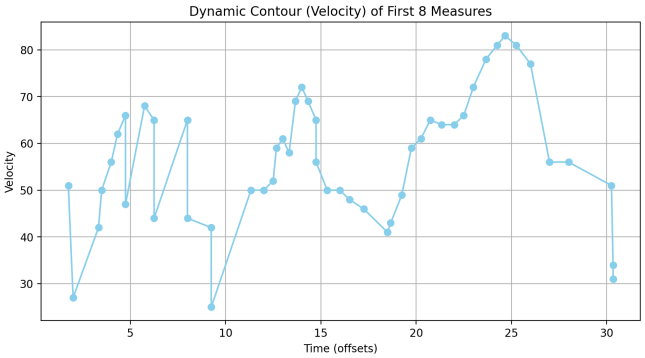

Figure 1: Chord ending of a Phrase

Timing also plays a crucial role in the unity of a phrase. Phrases often include variations in timing such as speeding up or slowing down towards the end, similar to physical gestures. This can involve subtle changes in the speed of individual notes and the stretching of time at the end of phrases, a practice known as rubato. Such timing flexibility helps in expressing the emotional content of the music and highlighting the phrase structure. [3] Similarly, Gingras et al. (2016) [4] explored how melodic predictability and expressive timing in performances influence perceived musical tension. They noted that variations in timing, driven by the predictability of melodic structures, significantly affect the expressive quality and emotional engagement of listeners, especially in semi-improvisatory genres like the unmeasured prelude. This underscores the complexity of musical phrases, where both dynamic and temporal elements play integral roles in crafting the overall expressive impact of a piece. As we listened in the piece, the timing variations follows this kind of contour.

Furthermore, phrases can also be characterized by the attributes of the symbolic score. For instance, they possess a melodic contour that distinguishes them from each other. In our example, phrases typically consist of an ascending phase followed by a descending one. However, there are various types of melodic contour in music and they vary between genres and geographical areas [5].

On the harmonic level, phrases tend to build up tension through the use of suspended chords or sevenths before resolving to the tonic or a more stable chord. This tension/resolution dynamic is mirrored on the rhythm level, where rhythms at the beginning of the phrase are often fast and syncopated, culminating in sustained whole notes or

a significant reduction in the number of notes per bar towards the end. For example, the final chord of one phrase can be illustrated as follows from our example [6]:

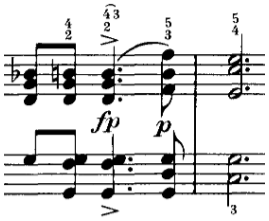

Figure 2: Chord ending of a Phrase

Alternatively, a complete silence may signal the end of the phrase:

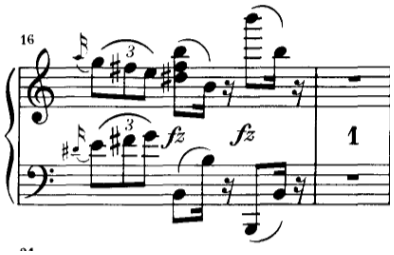

Figure 3: Silence ending of a Phrase

Or, the phrase may conclude with the silence of the right hand:

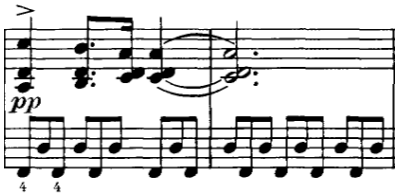

Figure 4: Left hand solo ending of a Phrase

Knowing whether phrases are accented at their beginning or their end rhythmically (hypermeter) is still a debate [7].

Dynamics play a crucial role in delineating phrases in scores, with crescendos and decrescendos following the melodic contour. It is also common to encounter low dynamics at the phrase's conclusion or to witness diminuendos or morendos. Conversely, there are instances where a phrase is punctuated by a loud final note, often characterized by an overall crescendo dynamic.

Tempo changes on the score level may also contribute to phrase demarcation, with ritardandos or fermatas employed to mark phrase endings, and accelerandos used to initiate them.

Finally, a key feature of the chosen score is the repeat symbol, as numerous phrases are repeated in our score [6].

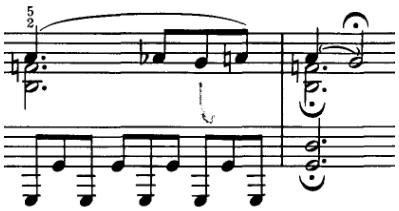

Figure 5: Fermata ending of a Phrase

2. ANNOTATIONS

There is several differences between phrase boundaries annotations [2] from hearing or reading the score. The most important one is around measures 18-23 and is explained below. We can also note that the silence of measure 17 or 83 would be part of the previous phrase when hearing it but from reading the score it is not really part of any phrase, it is more like a long boundary.

Measures 18 to 23: In the music score from measures 18 to 23, we can observe clear phrase boundaries, delineated by dynamic markings and rhythmic changes. Measures 18 to 21 form one phrase, which concludes with a gradual diminuendo, while the beginning of measure 22 marks the start of a new phrase, evidenced by changes in dynamics and the introduction of new rhythmic elements.

Although the phrase boundaries are distinct in the score, the continuity of performance attributes, particularly the connection between the left-hand rhythm in earlier measures (20-21) and the right-hand part in subsequent measures, makes the music sound smooth and continuous.

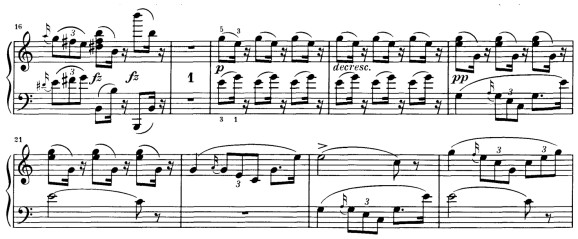

Figure 6: Measures 18-23

Measures 38 to 47: The phrase starting at the 38th measure could potentially be divided into three parts. The first two measures and the next two measures are almost identical, while the final three measures have a slightly different form. This phrase can thus be divided into three sections. However, in performance, it transitions smoothly and connects to the phrase starting at the 45th measure. While there is a clear distinction in the score, in performance, it flows smoothly, making it less likely to be perceived as separate sections.

Measures 48 to 52: An additional point of distinction can be observed between the 50th and 51st measures. A clear differentiation is that a new phrase begins at the 51st measure. However, there is ambiguity in the ending

[2] https://github.com/JorisMonnet/DM_Assignment3/blob/main/score_annotated.pdf

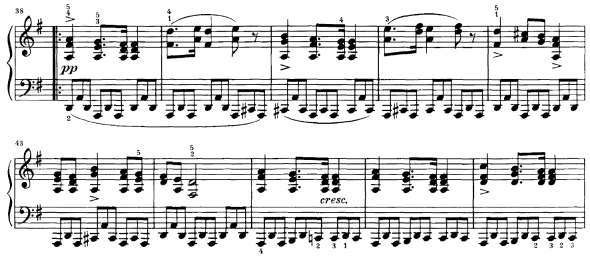

Figure 7: Measures 38-47

of the primary and secondary melodies in the 49th and 50th measures. The102 primary melody reaches its conclusion at the 49th measure, with the sound merely sustaining in the 50th measure. Conversely, the secondary melody continues with a similar rhythmic pattern in this measure. In terms of performance attributes, the phrase feels like it ends at the 49th measure, but in terms of score attributes, it appears to continue until the 50th measure. This is an example of ambiguity in the timing of the phrase's conclusion.

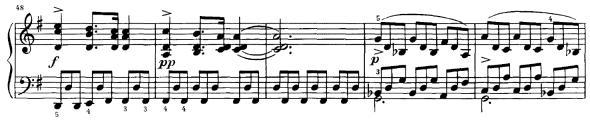

Figure 8: Measures 48-52

3. MODEL

3.1 C1: Model with performance attributes

We developed a simple phrase segmentation algorithm, `Model_P`, from two perspectives: changes in volume and timing. These features have been discussed as indicators of segmentation. By combining these two perspectives, the model can potentially identify phrase boundaries that might be missed when considering only one perspective. This approach contributes to the creation of a more robust and reliable model for phrase segmentation.

3.1.1 A model for distinguishing phrases based on changes in volume

In the previous chapters, we have identified that dynamics influence the distinction of musical phrases. Now, we will attempt to distinguish phrases based on dynamics. First, we will illustrate the overall changes in volume throughout the entire piece with a diagram. In Figure 9, we show the changes in volume throughout the entire piece, along with the segmentation points that can be inferred from these changes. For instance, it is believed that points where the volume significantly increases or decreases could indicate the boundaries of phrases.

From this perspective, the model can be constructed as follows:

1. Capture the changes in dynamics between adjacent notes.

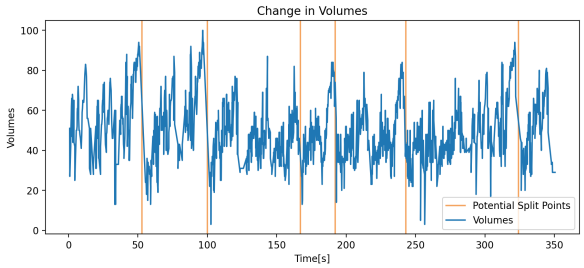

Figure 9: Change in Volumes

2. To eliminate noise, square the changes in dynamics obtained in the previous step.

3. Normalize each squared change in dynamics by dividing it by the maximum squared change in dynamics.

This approach allows for the removal of noise from the volume changes between adjacent notes, enabling the extraction of significant values within a range of 0 to 1. Subsequently, by setting an appropriate threshold, we can identify points with substantial changes.

The squared and scaled changes in volume between adjacent notes are shown in Figure 17. While most values have become very small except for the prominent ones, the noise has been completely eliminated. Values with even a slight possibility remain around 0.1.

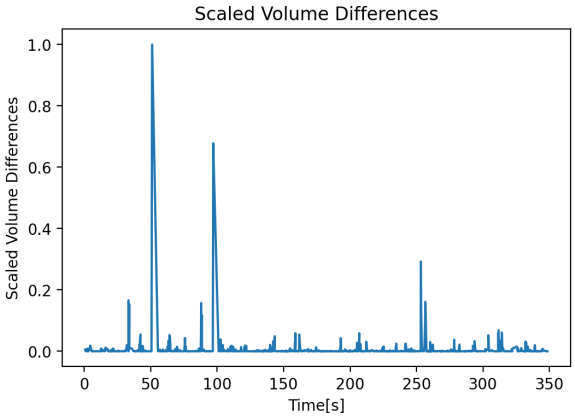

Figure 10: Squared and Scaled Differences in Volumes between adjacent notes

3.1.2 A model for distinguishing phrase based on changes of timings

We also identified that tempo changes can signify boundaries between musical phrases. Specifically, a decrease in tempo within one measure followed by an increase in the next measure often indicates a phrase boundary, symbolizing a ritardando or fermata before a new phrase begins.

Given this understanding, we implemented a function, get phrase boundaries, which analyzes tempo changes to detect phrase boundaries in musical pieces. The function processes a tempo map, identifying measures where the tempo decreases followed by an increase, thus marking potential phrase boundaries. The tempos are computed as a

ratio between the tempo of the performed piece and the unperformed one using the annotations files. This has been possible by using part of our Assignment 1. We assume that a phrase is starting on a downbeat.

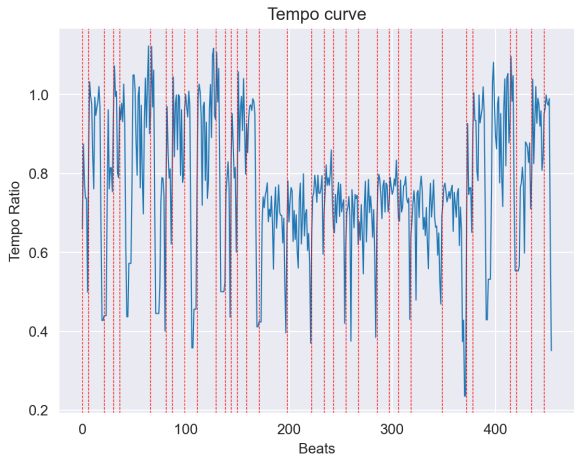

Figure 11: Tempo ratios with Phrase boundaries

3.1.3 A full model on performance attributes

The two models presented in the previous sections are subsequently merged, each providing a list of timestamps for the phrase boundaries. We then check for concordance between the two sets of boundaries within a small margin of error. The downbeat position information from the timing model is used to identify the closest downbeat for each boundary.

Additionally, if one model fails to identify boundaries at the extremes (as seen with the velocity model in our example), we retain the boundaries detected by the other model. This approach ensures that both models complement and verify each other without discarding valuable information. Finally, we display the measure numbers corresponding to these boundaries, providing a clear overview of the identified phrase divisions.

3.2 C2: Results of Performance attributes based model on Moment Musical n°1

The results are surprisingly good for this model, which is not using any kind of repetitions (see paragraph below) to segment phrases. It outputs phrase boundaries we had previously identified, such as in measure 66 where it finds the fermata:

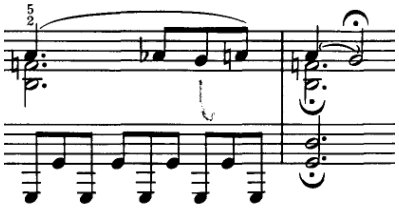

Figure 12: Measure 66

Or in measure 8 where it finds the final chord of the phrase:

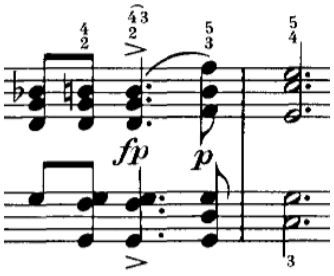

Figure 13: Measures 7-8

However, there are some discrepancies. Some boundaries are incorrect, and others are correct but segment subphrases instead of the larger phrases we annotated in Task B. For example, both times this measure is played, the model segments phrases between measures 2 and 3:

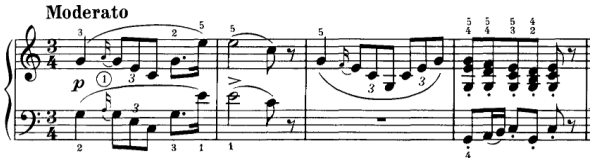

Figure 14: Measures 1-4

This segmentation is valid because there is a silence and a change of tempo in the performance, but it operates at a lower level than our annotations.

3.3 C3: Repetition Analysis Model

This model employs repetitive patterns to delineate phrase boundaries, focusing on both pitch and rhythm elements. Rather than relying solely on exact pitch matches, it adopts a more flexible approach using intervals to identify recurring melodic contours. This strategy allows for a comprehensive analysis that captures the structural essence of the music, accommodating variations in pitch while preserving the underlying patterns.

3.3.1 Pattern Extraction

The "find repeating sequences" function identifies recurring patterns in sequences, encompassing pitch (bass in chord sequences), intervals (derived from preceding bass notes), and durations for rhythm analysis.

Data Extraction: The model initially extracts essential data directly from the MIDI file, including measure numbers, offsets, pitch, and duration.

Pitch Intervals Calculation: Pitch intervals are computed by determining the difference in MIDI note numbers between consecutive bass notes. For instance, if a sequence comprises C4 (MIDI number 60) followed by E4 (MIDI number 64), the interval would be 4 semitones (64 - 60 = 4).

3.3.2 *Pattern Filtering*

The "filter sub patterns" function refines identified patterns by ensuring retention of only unique and independently significant ones.

Sub pattern Identification: Each identified pattern is compared against others to ascertain if it is a subset of another. Subsequent patterns found within larger ones across the dataset are deemed sub patterns.

Filtering Mechanism: Sub patterns are discarded as they fail to provide additional information beyond what is already captured by larger patterns. This step optimizes the analysis by focusing on the most musically significant patterns.

3.3.3 *Boundaries Detection*

For each pattern, only the first measure of occurrence is retained, assuming that phrases commence on the downbeat. Boundary detection integrates root, intervals, and durations patterns concurrently to identify phrases. Boundaries that are too close to each other (less than 2 measures) are filtered out to ensure accurate phrase delineation.

3.3.4 *Results on selected piece*

In Moment Musical No. 1, the model successfully identifies several phrases that were manually annotated during task B. However, it tends to output an excessive number of phrases, detecting segments that are slightly shorter than expected (for example, 1 and a half measures) or sub phrases. This results in an output of 28 phrase boundaries instead of around 15.

4. RESULTS

4.1 C4: Results on Larger Dataset with Model P

We ran Model P on the entire dataset, leaving only five pieces unanalyzed, and generated the following graph:

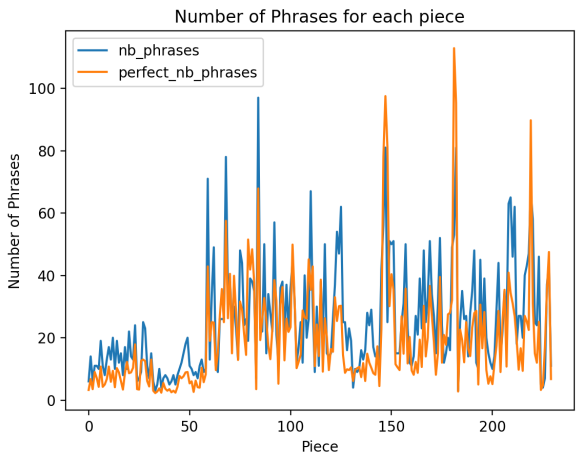

Figure 15: C4 Output Model P

In blue, the graph shows the number of phrases detected by our model. In orange, it shows the expected number of phrases if each phrase were 8 bars long, which is the most common phrase length. Our algorithm generally identifies more phrases than this average, and sometimes it identifies significantly more. This discrepancy can be attributed to various factors, but as explained in section C1, it often arises because the model detects sub-phrases. When considering phrases of 4 bars in length, our algorithm's output is closer to the "ideal" number of phrases. With 8 bars-long phrases, our model outputs on average on the whole dataset 1.53 times the "ideal" number of phrases.

4.2 C4: Results on larger dataset for the second model

We opted to execute the second model on a reduced dataset due to its computationally demanding nature. Specifically, we applied it to the Bach compositions from the asap dataset. On average, the model yields 1.83 times more phrase boundaries than an "ideal" phrase length of 8 bars. As depicted in the graph, certain pieces exhibit numerous detected phrases. However, when selecting pieces randomly, the identified phrases generally align closely with those a human annotator would mark, although some are subphrases, akin to observations made with the previous model.

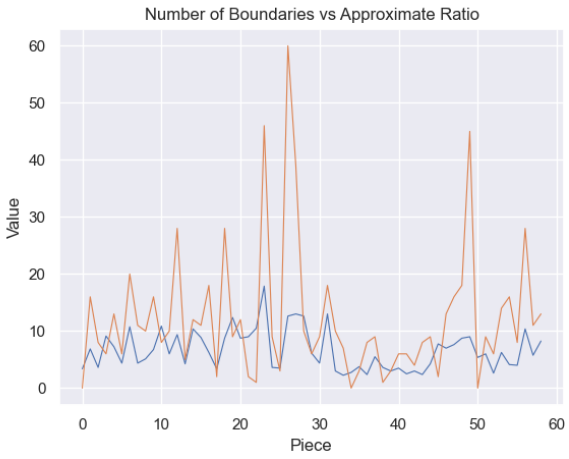

Figure 16: C4 Output 2nd Model

4.3 C4: Analysis of 1 piece: Bach - Fugue BWV 848

Our initial model identifies 14 phrases, while the second model detects 16 phrases in this 54-measure piece. Upon closer examination of the measure numbers, it becomes apparent that the models identify a phrase every two measures at the beginning, likely due to the frequent changes in dynamics and timings.

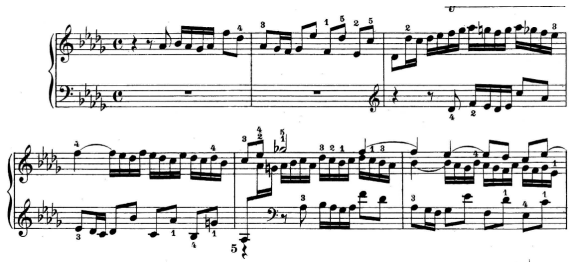

Figure 17: First measures of Bach bwv 848

While this segmentation accurately captures the sub phrases present at the start, it deviates from manual annotation, which typically aims for a higher-level segmentation. This discrepancy highlights an area for improvement in our models: they tend to produce an excessive number of boundaries due to their tendency to segment at a lower level.

Interestingly, the first model performs relatively better in identifying phrase boundaries, particularly in cases where the piece contains fewer repetitions compared to Moment Musical No. 1. In this piece, repetitions are not as pronounced at the phrase level but manifest more rhythmically at the measure level. Our selection of this piece was deliberate, aiming to assess whether the second model could effectively handle a piece without significant repetitions of 8 or 21 measures. The results demonstrate that the second model is indeed capable of working with such pieces, suggesting its potential utility in enhancing model P. However, both models exhibit a bias towards segmenting at a lower level than the human-perceived phrase boundaries.

Despite these limitations, it is intriguing to observe how our models can detect phrases with relatively few parameters and how they perform on new pieces without extensive fine-tuning. This exploration underscores the ongoing quest to refine computational models for musical phrase segmentation, aiming for alignments closer to human annotation standards.

5. DISCUSSION

There is still room for improvement, as some attributes have not been implemented in our models. For performance attributes, a major addition would be articulation. For example, detecting transitions between staccato and legato phrases could enhance our model. Additionally, timing and velocity attributes could be further refined to handle different kinds of patterns commonly used for phrase segmentation.

For score attributes, the model could be improved by allowing a margin of error for repetitions of pitches or rhythms. Detecting changes in rhythm or articulation would also be beneficial, as these are important indicators of phrase boundaries. Finally, incorporating melodic contour analysis by identifying patterns specific to certain genres of music could further enhance phrase detection.

6. CONCLUSION

In this paper, we explored the interpretation and perception of musical phrases within a digital musicology context, focusing on **Moment Musical n°1** by Schubert. We developed a computational model to identify phrase boundaries based on MIDI velocity, dynamic contours, and rhythmic patterns, and tested its effectiveness on a larger dataset. Our findings highlight the complex interplay of dynamics, timing, and melodic contour in shaping musical phrases.

Our model showed promising results in detecting phrase boundaries, correctly identifying key points such as fermatas and final chords. However, it also revealed some discrepancies, particularly in sub-phrase segmentation and boundary accuracy. This indicates that while our model captures significant aspects of phrase structure, there is still room for improvement.

7. AUTHOR CONTRIBUTION

Task A1: Yiwei Liu
Task A2: Joris Monnet
Task B1: All
Task B2: Yutaka Osaki, Xingyu Pan, and Joris Monnet
Task C1: Yutaka Osaki and Joris Monnet
Task C2: Yutaka Osaki and Joris Monnet
Task C3: Yiwei Liu and Xingyu Pan and Joris Monnet
Task C4: Joris Monnet
Report: All
Code Cleaning: All
Notebook: All

8. REFERENCES

[1] N. P. McAngus Todd, "The dynamics of dynamics: A model of musical expression," *The Journal of the Acoustical Society of America*, vol. 91, no. 6, pp. 3540–3550, 1992.

[2] K. N. Olsen, C. J. Stevens, R. T. Dean, and F. Bailes, "Continuous loudness response to acoustic intensity dynamics in melodies: Effects of melodic contour, tempo, and tonality," *Acta Psychologica*, vol. 149, pp. 117–128, 2014, including Special section articles of Temporal Processing Within and Across Senses - Part-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001691814000821

[3] H. Shaffer, "Timing in musical performance a," *Annals of the New York Academy of Sciences*, vol. 423, no. 1, pp. 420–428, 1984.

[4] G. Ren, D. Headlam, S. Roessner, J. Lundberg, and M. F. Bocko, "Extracting heterogeneous structured features from music performance," *The Journal of the Acoustical Society of America*, vol. 130, no. 4_Supplement, pp. 2431–2431, 2011.

[5] B. Cornelissen, W. Zuidema, and J. A. Burgoyne, "Cosine Contours: a Multipurpose Representation for Melodies," in *Proceedings of the 22nd International*

Society for Music Information Retrieval Conference. ISMIR, Oct. 2021, pp. 135–142. [Online]. Available: https://doi.org/10.5281/zenodo.5624531

[6] "Imslp complete original score," https://ks15.imslp.org/files/imglnks/usimg/e/ed/IMSLP38719-PMLP02066-Schubert-6Moments-Musicaux-D780-ed-Henle.pdf, last accessed the 09/05/2024.

[7] S. Ng, "End-Accented Sentences: Towards a Theory of Phrase-Rhythmic Progression," *Music Theory Spectrum*, vol. 43, no. 1, pp. 43–73, 01 2021. [Online]. Available: https://doi.org/10.1093/mts/mtaa018