A multi-modal platform for semantic music analysis: visualizing audio- and score-based tension

Dorien Herremans
School of Electronic Engineering and Computer Science
Queen Mary University of London
Mile End Road, E1 4NS, London, UK
Email: d.herremans@qmul.ac.uk

Ching-Hua Chuan
School of Computing
University of North Florida
1 UNF Drive, Jacksonville, USA
Email: c.chuan@unf.edu

***Abstract*—Musicologists, music cognition scientists and others have long studied music in all of its facets. During the last few decades, research in both score and audio technology has opened the doors for automated, or (in many cases) semi-automated analysis. There remains a big gap, however, between the field of audio (performance) and score-based systems. In this research, we propose a web-based Interactive system for Multi-modal Music Analysis (IMMA), that provides musicologists with an intuitive interface for a joint analysis of performance and score. As an initial use-case, we implemented a tension analysis module in the system. Tension is a semantic characteristic of music that directly shapes the music experience and thus forms a crucial topic for researchers in musicology and music cognition. The module includes methods for calculating tonal tension (from the score) and timbral tension (from the performance). An audio-to-score alignment algorithm based on dynamic time warping was implemented to automate the synchronization between the audio and score analysis. The resulting system was tested on three performances (violin, flute, and guitar) of Paganini's Caprice No. 24 and four piano performances of Beethoven's Moonlight Sonata. We statistically analyzed the results of tonal and timbral tension and found correlations between them. A clustering algorithm was implemented to find segments of music (both within and between performances) with similar shape in their tension curve. These similar segments are visualized in IMMA. By displaying selected audio and score characteristics together with musical score following in sync with the performance playback, IMMA offers a user-friendly intuitive interface to bridge the gap between audio and score analysis.**

***Index Terms*—Multimodal system, music analysis, tension, online interface, music representation**

I. Introduction

Extracting low-level features such as chroma vectors and Mel-frequency spectral coefficients from audio recordings has been the driving force for Music Information Retrieval (MIR). The success of various MIR tasks, including music recommendation and playlist generation, requires the analysis of audio as a fundamental step. More recently, researchers in MIR started to examine the semantic meaning of these low-level acoustic features. The most common "multi-modal" approach in MIR is to study the relation between these low-level acoustic features and high-level features labeled by music experts or casual listeners. For example, Schedl and et al. [1] created an informative dataset of popular music by considering acoustic features, user-generated tags (free-style short textual segments) from Last.fm, expert-annotated tags (genre and mood), and editorial meta-data such as album and artist name. Similarly, Wang and et al. [2] combined user-generated tags and acoustic low-level features as ontology classes for music recommendation. Saari and Eerola [3] also used social tags from Last.fm to study the semantic meaning of mood in music.

The problem with the above-mentioned approach is that it is not truly multi-modal. Such an approach mainly focuses on acoustic features, and only aims to create another level of representation that groups certain acoustic features together using tags as categories. In addition, using user-generated tags to study semantic meanings in music is convenient but superficial: one tag such as "coffee house" for a song does not describe the musical nuances in the composition or recording. To truly understand semantically meaningful concepts such as tension, relaxation, and closure, it is necessary to consider the manner in which music theorists and musicologists study music, i.e., to study the score.

Several software tools exist for music editing and semantic music analysis, but none of them focus on the features that we aim for in this paper. For example, Sonic Visualizer [4] allows users to annotate and visualize acoustic characteristics in audio recordings. Through the use of Vamp plug-ins[1], the software can automatically perform analyses such as identifying the fundamental frequency of notes and extracting beat, tempo and chords from a music recording. However, displaying or analyzing the musical score is not currently supported in the software, nor does the system work online. In contrast, MuseScore[2] provides powerful functions to create and annotate a music score, but it does not support semantic music analysis on the score nor audio level.

We propose a web-based Interactive Multi-modal Music Analysis system (IMMA) that provides a multi-modal interface that unifies audio- and score-based musicological analysis. In this paper, we focus on tension as an initial use-case for semantic music analysis. "Music, particularly music in the Western tonal-harmonic style, is often described in terms of patterns of tension and relaxation" [5]. Modeling musical tension has long captivated the attention of researchers in music theory and psychology. Empirical studies indicate that many aspects of

[1] http://www.vamp-plugins.org/

[2] https://musescore.org/

Preprint published in the Proceedings of 11th International Conference on Semantic Computing (IEEE ICSC), San Diego, US, Feb. 2017

music including melody [6], harmony [6], [7], tempo [8], and phrase structures [9] are highly correlated to perceived tension, among both musicians as well as non-musicians [10]. Ratings of tension have also been found to correlate with ratings of physiological measures [9]. In addition, studies have found that the effect of tension-relaxation patterns is a key component in aesthetic experience and emotions [11].

In order to design a system that supports musical tension analysis, it is necessary to understand the research methodologies and needs of researchers working on this topic. Researchers who study melodic and harmonic tension usually provide detailed analyses on segments of one or more pieces, mapping the predicted tension from their model, or the tension ratings from the participants of a listening experiment, to the score note-by-note. In order to properly assess tension in a listening experiment, the ideal stimulus is a *performance* of the piece by a musician. This provides a more authentic music experience, yet, it also requires control of potentially interrelated variables, i.e., a profound understanding of both the composition (score) and performance (audio) is required to examine results of a listening experiment. For example, in Krumhansl's experiment [5], the stimulus consisted of a performance of Mozart's piano sonata K. 282, played by a musician and recorded into MIDI format for playback, with a detailed analysis on the performance by Palmer [12]. More recently, Farbood and Price [13] explored the contribution of timbre to musical tension by using artificially synthesizing sounds with two states of spectral characteristics as stimuli in their experiment. The authors also indicated the future research direction: "The next step is to explore precisely how these features covary in order to model how dynamic timbral changes influence tension perception. Additional experiments using more complex stimuli—particularly musical stimuli where other musical features influencing tension such as harmony and melodic contour are involved—are the next directions to explore.". This statement confirms the need for a multi-modal system to examine results from both score- and audio-based analyses.

A review of the literature thus pinpoints the important functions that IMMA offers: 1) representing the score with a graph that shows multiple semantic score-based characteristics such as tonal tension; 2) aligning audio recordings to the score to show how audio-based timbral features change over time in relation to the score; 3) providing functions for cross-correlation analysis between multi-modal features; and 4) retrieving segments with similar tension patterns for detailed analysis. To demonstrate the capability of the system, we study four piano performances of Beethoven's Moonlight sonata and three performances, with different instruments, of Paganini's Caprice No. 24 in this paper. For each of these performances, the audio recording is first aligned to the score using a dynamic time warping (DTW) algorithm (Section II). Three tonal tension models [14] and five timbral features related to tension [13] are implemented to extract tension data from both the score and the audio (Sections III-A and III-B). The relationship between tonal and timbral tension is analyzed statistically and visualized in plots and with ribbons over a score for easy interpretation (Section III-C). A clustering algorithm is performed to compare performances based on similar tension patterns (Section III-D). IMMA interactively visualizes these analyses and is available as a web-based application for easy accessibility at http://dorienherremans.com/imma.

II. Audio-to-score alignment

In order to create a multi-modal interface for semantic music analysis based on audio as well as score, the first task is to synchronize the audio recording with the score, i.e., to identify when each note in the score is played in the audio performance. This task is called audio-to-score alignment in MIR. Audio-to-score alignment has been extensively studied and remains a popular topic in the MIR community [15]–[17]. We chose to adopt and modify the approach as described in [18]. This algorithm uses DTW to calculate the distance between two sequences and to extract the optimal alignment path based on the distance. DTW has been used in all submissions for MIREX 2014 and 2015 score following competitions[3]. In this study, one sequence consists of the chroma representation over time for the audio recording, and the other represents the chroma of the synthesized audio from the MIDI file that represents the score. Both the audio recording and the synthesized audio are analyzed using fast Fourier transform with half-overlapped windows to extract a chroma vector roughly every 185 milliseconds.

To evaluate the alignment result, we manually annotated the onsets using Sonic Visualiser[4] for the first 60 seconds of each performance and compared the annotated onsets with the aligned time. We adopted the evaluation metrics from MIREX score following task which are based on piecewise precision rate, i.e., the average percentage of detected notes, which are defined as the detected onset within a tolerate threshold of the actual onset. The results are shown in Table I. Readers can also assess the alignment results by listening to the synchronized examples, the audio recording on the left channel and the aligned re-synthesized audio on the right, at *http://dorienherremans.com/imma.*

In the future, we will incorporate the active learning algorithm described in [19] so that IMMA can improve the accuracy of audio-to-score alignment by augmenting the automated alignment with manual corrections on the most uncertain note onsets. The alignment algorithm implemented in IMMA allows us to visualize semantic characteristics of both audio and score in a synchronized way.

threshold (ms.)	250	500	750	1000	1250	1500
precision rate	0.69	0.89	0.96	0.98	0.98	0.99

TABLE I: Piecewise precision rate for the first 60 seconds of the seven audio recordings studied in this paper.

[3] http://www.music-ir.org/mirex/wiki/2014:Real-time_Audio_to_Score_Alignment_(a.k.a._Score_Following)_Results

[4] http://www.sonicvisualiser.org/

III. Synchronized musicological characteristics

In this section, we demonstrate how the proposed system can be used as a tool for multi-modal musicological analysis by analyzing tension characteristics calculated from both audio and score files.

Farbood [8] describes increasing musical tension as “a feeling of rising intensity or impending climax, while decreasing tension can be described as a feeling of relaxation or resolution”. Tension is a complex, composite characteristic that is not easy to quantify. Musicologists therefore often look at different aspects of tension when studying a piece or performance. In this paper we will discuss three aspects of tonal tension based on [14] and align them to different timbral characteristics [13] extracted from the audio signal. The results and benefits of the alignment methods for analyzing tension are discussed based on one of Beethoven’s most popular piano pieces, Sonata No. 14 in C♯ minor “Quasi una fantasia”, Op. 27, No. 2 (otherwise known as the Moonlight Sonata), and Paganini’s Caprice No. 24, performed with three instruments: violin, flute, and guitar. We also demonstrate how IMMA can cluster and visualize segments based on semantic similarity. Finally, the implementation details of the system are discussed.

A. Tonal tension ribbons based on score

Different aspects of tonal tension were captured from a musical score with a model for tonal tension [14] based on the spiral array [20]. The spiral array is a three dimensional representation of pitch classes, chords and keys. Each pitch class is represented as spatial coordinates along a helix. The spiral array is constructed in such a way that close tonal relationships are mirrored by their close proximity in the array [21]. This concept is illustrated in Figure 1 in which a C-major chord is drawn in the array (in blue).

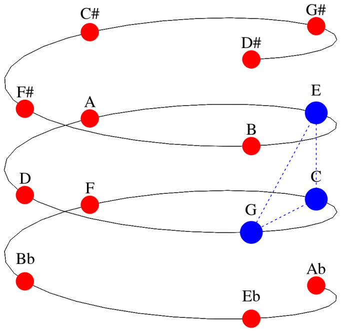


Fig. 1: The helix of pitch classes in the spiral array [20]

In [14], Herremans and Chew present three methods for quantifying aspects of tonal tension based on the spiral array. In order to do so, the piece is first divided into equal length segments, which form a cloud of points in the spiral array.

Based on this cloud of notes, the first aspect of tonal tension captures the dispersion of notes in tonal space and is calculated as the cloud diameter. Figure 2 illustrates that a tonally consistent chord has a small cloud diameter. The first chord, which consists of the notes D and G, has a very small diameter in the spiral spiral array, as can be seen in Figure 1. The third chord, which consists of the notes D, G and E♯, is tonally very dispersed and thus has a large cloud diameter.

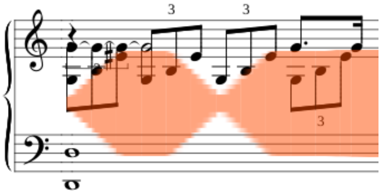


Fig. 2: Cloud diameter ribbon on a fragment from Beethoven’s Moonlight Sonata.

Cloud momentum, a second aspect of tonal tension, captures the movement of subsequent clouds (i.e. chords) in the spiral array. The tonal movement in the opening bar of the Moonlight sonata is displayed in Figure 3. As long as there is an arpeggiation over the same chord, there is no change in cloud momentum, but when the chord changes on the third beat, the cloud momentum ribbon clearly indicates a movement in tonal space.

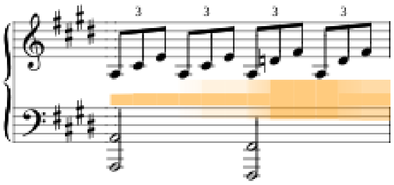


Fig. 3: Cloud momentum ribbon on a fragment from Beethoven’s Moonlight Sonata.

Finally, tensile strain measures how far the tonal center of each cloud is removed from the global key. Figure 4 illustrates how the cloud momentum ribbon grows bigger when there is a movement from notes predominantly belonging to A minor (the global key) to G♯ and F♯.

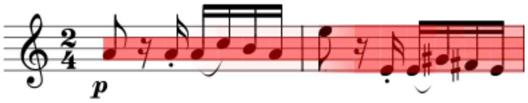


Fig. 4: Tensile strain ribbon on the first two bars of Paganini’s Caprice No 24.

These three methods are implemented in a system that visualizes the results as tension ribbons over the musical score, allowing for easy interpretation [14]. This system is integrated in IMMA, which ports the results into the interactive score characteristics plot (see Figure 7).

B. Timbral tension based on audio

The five features used to capture timbral tension in this paper are based on [13]. These features include loudness, roughness, flatness, spectral centroid, and spectral spread/deviation. Loudness is measured via the root-mean-square of the audio wave amplitude. Roughness measures the sensory dissonance by calculating the ratio between pairs of peaks in the frequency spectrum. Flatness shows how smooth the spectrum distribution is as the ratio between the geometric mean and the arithmetic mean. Spectral centroid and spread calculate the mean and standard deviation of the spectrum. Each of these features has shown to contribute to perceived tension, however, they have yet to be integrated in one comprehensive model [13]. These features were extracted using MIRToolbox [22] with half-overlapped windows, similar to the windowing approach used for the alignment process. Based on the alignment result, the average per window is calculated for each timbral feature. This value is then mapped to the aligned onsets, so that it can be synchronized to and compared with tonal tension.

In addition to these five timbral features, the system estimates tempo variations of the audio performance based on the alignment result. In order to reduce the impact caused by alignment errors on the estimation, we calculated the local alignment cost at each aligned point and excluded the points where the cost is above 95% threshold for tempo estimation.

C. Synchronizing tension based on score and audio

The alignment of tonal tension ribbons and audio-based timbral tension features allows us to examine how the different aspects of tension correlate over different performances of the same piece. The analysis results for four performances of the Moonlight sonata are displayed in Table II. Table III shows correlation results of three performances of Paganini's Caprice No 24, each with a different instrument.

A correlation analysis was performed on the data, with a window size of one quarter note. Since tension is typically cyclic throughout a piece, there is autocorrelation within each of the tension features, which influences the interpretability of cross-correlation [23]. We therefore fitted an Arima model each of the characteristics and used this to prewhiten the data, so that the trend is removed. The resulting cross-correlation values calculated with the software package R[5] are displayed in Tables II and III.

When interpreting these results, we should keep in mind that tension is a composite characteristic. The different characteristics described in this paper capture different aspects, and may therefore not always be correlated. Yet as a first analysis, it can give us insight into strongly correlated characteristics. Examples of highly correlated audio and score-based tension characteristics are shown in Figure 5.

Audio based			Score based	
		Diameter	Momentum	Tensile strain
Loudness	A1	0.187 (0)	0.309 (1)	N/A
	A2	0.201 (-1)	0.252 (0)	0.117 (4)
	A3	0.307 (-1)	0.253 (1)	0.124 (3)
	A4	0.239 (-2)	0.114 (1)	N/A
Roughness	A1	0.153 (0)	0.332 (1)	-0.131 (-2)
	A2	0.181 (0)	0.277 (0)	0.108 (3)
	A3	0.302 (-1)	0.283 (1)	N/A
	A4	0.198 (0)	N/A	-0.121 (0)
Flatness	A1	-0.277 (0)	-0.368 (-1)	0.137 (-4)
	A2	-0.201 (-2)	-0.213 (1)	N/A
	A3	-0.219 (0)	-0.221 (1)	N/A
	A4	-0.285 (-1)	-0.164 (1)	N/A
Centroids	A1	-0.207 (0)	-0.344 (1)	0.132 (-4)
	A2	N/A	-0.160 (1)	N/A
	A3	-0.117 (2)	-0.198(1)	N/A
	A4	N/A	-0.198 (1)	N/A
Spread	A1	-0.218 (0)	-0.367 (1)	0.120 (-4)
	A2	-0.175 (2)	-0.213 (1)	N/A
	A3	-0.257 (0)	-0.216 (1)	N/A
	A4	-0.327 (0)	-0.187 (1)	N/A

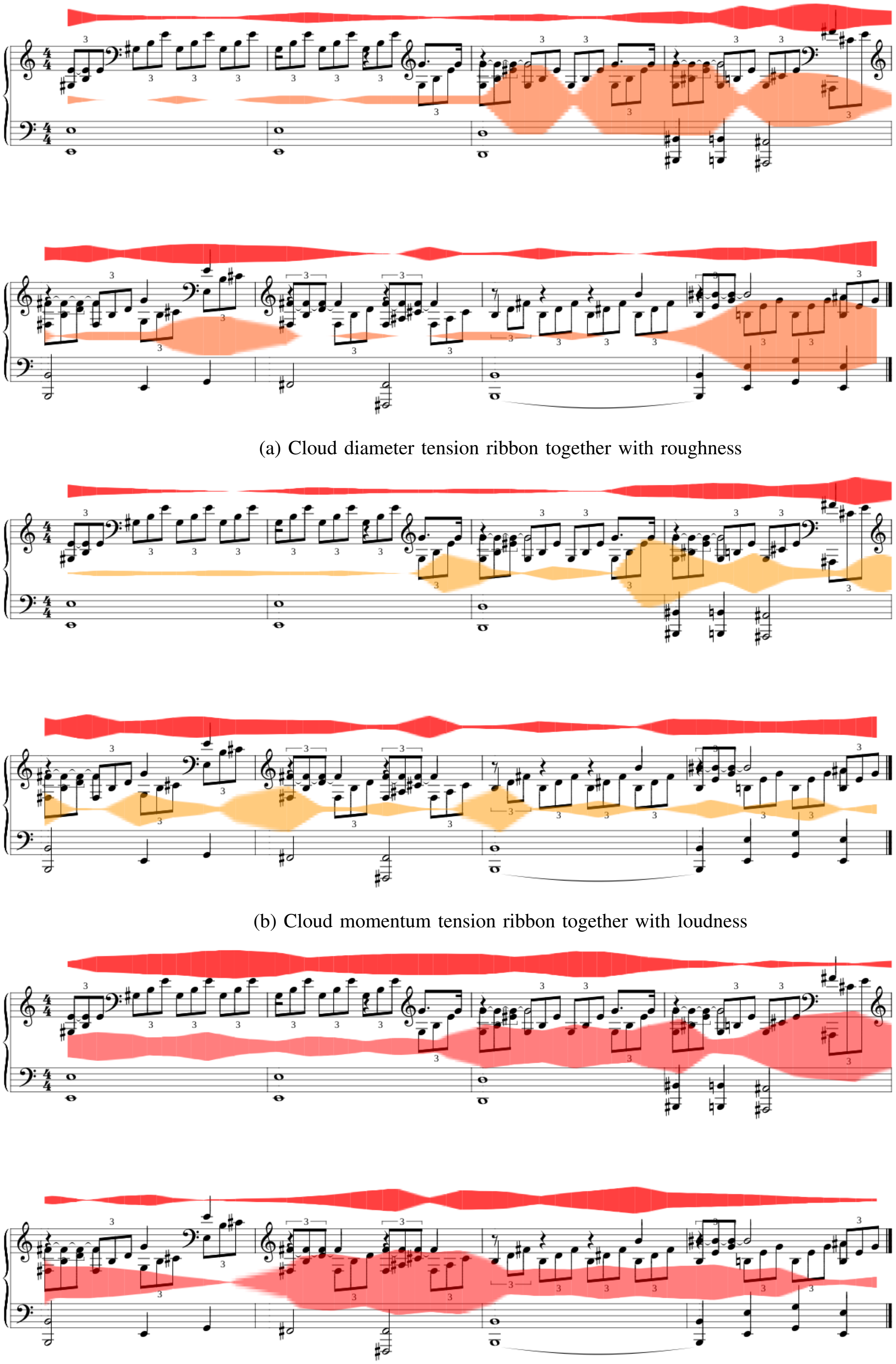
TABLE II: Highest significant cross-correlation coefficient (after prewhitening) together with its lag (1 unit = 1 window) between tonal tension characteristics and aligned timbral tension features (N/A = no significant correlation) based on Beethoven's Moonlight Sonata. The performances are by Evgeny Kissin (A1), Wilhelm Kempff (A2), Arthur Rubinstein (A3) and Tiffany Poon (A4).

The analysis results of the Moonlight sonata in Table II show that there is a consistent significant correlation of roughness/loudness with cloud diameter/cloud momentum for most of the performances. The correlation between tensile strain and the timbral features is not significant, except in the case of the performance of Evgeny Kissin, for which tensile strain is positively correlated with flatness, centroids and spread. The negative correlation between, for instance, cloud diameter and flatness, confirms that tension is a complex concept that consists of an interplay of different aspects. In the case of Beethoven's sonata, certain tension characteristics such as flatness and cloud diameter seem to have an interchanging dynamic.

The proposed system does not only allow us to study the effect of performance on tension, but also the influence of instrumentation. Table III shows three performances, each with a different instrument, performing Paganini's Caprice No 24. In contrast to the previously discussed piece, loudness is not correlated with the cloud diameter. It is, however, correlated with cloud momentum and (in some cases), the tensile strain. It is to be expected that greater variations exist in the size and direction of the correlations, since instrumentation has an important effect on timbral features such as roughness. Different instruments manipulate distinctive aspects of timbre, thus allowing them to express tension in different ways, as is confirmed by the correlation results.

We have analyzed the correlation of the tension characteristics throughout the entire piece. In the next section, we discuss an example of how the proposed system can identify smaller musical fragments within a piece that have similar properties in tension features.

[5] r-project.org



(a) Cloud diameter tension ribbon together with roughness

(b) Cloud momentum tension ribbon together with loudness

(c) Tensile strain tension ribbon together with flatness

Fig. 5: Selected score based and timbral tension characteristics based on Kissin's performance of Beethoven's Sonata No. 14 in C♯ minor, Op. 27, No. 2, bars 9–16. The score characteristics are displayed over the score, the audio features are depicted above the score.

Audio based		Score based		
		Diameter	Momentum	Tensile strain
Loudness	Violin	N/A	0.097 (1)	N/A
	Guitar	N/A	0.199 (-1)	0.174 (-3)
	Flute	N/A	0.424 (0)	0.354 (0)
Roughness	Violin	0.123 (-2)	N/A	-0.120 (-2)
	Guitar	0.173 (-1)	-0.106 (0)	N/A
	Flute	-0.131 (-4)	0.511 (-4)	0.395 (-3)
Flatness	Violin	0.151 (4)	-0.258 (3)	-130 (-4)
	Guitar	0.440 (3)	-0.583 (-2)	-0.425 (3)
	Flute	0.174 (2)	0.096 (0)	N/A
Centroids	Violin	0.125 (0)	-0.209 (2)	-0.103 (-4)
	Guitar	0.368 (3)	-0.402 (4)	-0.261 (2)
	Flute	0.190 (2)	0.243 (0)	0.170 (0)
Spread	Violin	N/A	-0.187 (2)	N/A
	Guitar	0.192 (-2)	-0.311 (0)	-0.247 (-4)
	Flute	0.185 (-3)	-0.200 (-1)	-0.201 (-4)

TABLE III: Highest significant cross-correlation coefficient (after prewhitening) together with its lag (1 unit = 1 window) between tonal tension characteristics and aligned timbral tension features (N/A = no significant correlation) for Paganini's Caprice No 24. The performances are by Julia Fisher (violin), Eliot Fisk (guitar), and Janos Balint (flute).

D. Clustering segments based on semantic similarity for performance analysis

In this section, we demonstrate how segments of a score or a performance can be clustered based on similar (tension) characteristics and visualized over the aligned score. Traditionally, the (audio) sources of a performance are analyzed by connecting the dynamics, such as loudness/tempo variations and articulation, to the score. The tension-based performance analysis included in IMMA provides an opportunity to link musical performance strategies with musical *segments* that have specific tension-relaxation patterns.

The performance analysis process starts with a "ribbon cutting" process which segments the score into segments for each tension characteristic (ribbon) by cutting the ribbon at the thinner points (local minima). Each ribbon segment is then clustered into groups based on its shape. The shape of each ribbon segment is described by its average height, the maximal height, width, and the angles of left and right slopes. K-means clustering is then used to encode each ribbon segment by the centroid of its group. Finally, the sequence of ribbon segments are represented using n-gram models to study the frequency of occurrence of each n-gram pattern and to retrieve the parts of the score that share similar tonal tension patterns.

Figure 6 shows an example of a performance analysis based on cloud diameter tension on the four performances of Beethoven's Sonata No. 14 in C♯ minor. In this example, only the height and width of a ribbon segment are considered in k-means clustering (k=5) and the sequences of ribbons are represented as tri-gram patterns. The tri-gram tension sequence shown in Figure 6 occurred four times in the score, highlighted in rectangles, at measures 25, 47, 55, and 58. The graphs below the scores show the loudness and tempo variations in the four performances for the identified tension sequence at measures 25 (sequence no. 3) and 47 (sequence no. 4). Although sequences no. 3 and 4 do not share exactly the same notes, similar trends in loudness variations (e.g. becoming louder towards the end of the sequence) are observed. Some similarity can also be spotted in tempo variations. However, it is not as consistent as loudness.

E. IMMA as a web application

The IMMA system is implemented as an interactive application, see Figure 7. Its interface allows for easy interpreting of a performance on both the score and audio level. This multi-modal system displays aligned musical analysis results of both audio and score. In this paper we have elected to focus on tension as an initial musical characteristic, yet, in future versions modules for other types of analysis will be added. The implementation details of IMMA include:

1) VexFlow API: The user can upload a score in musicXML format, an open format designed for easy interchanging of scores [24]. This file is then parsed with the VexFlow MusicXML plugin[6], and displayed as a score by the VexFlow API[7]. VexFlow is a rendering engine, built in JavaScript with the jQuery API, that displays a score on an HTML5 canvas with scalable vector graphics support.

2) Multi-modal music analysis: The IMMA interface allows users to playback an mp3 performance. A score following plugin was written for VexFlow that displays a colored box over the current bar of the score, synced with the audio. The music analysis of both the score and audio are displayed using Flot Charts[8], a JavaScript library for displaying interactive plots [25]. A moving crosshair over the plots is synced with the audio playback, allowing for an easy and user friendly interface for multi-modal music analysis (see Figure 7). The similar semantic fragments, as described in the previous section, are visualized as colored boxes over the plots. We decided to implement the music analysis results as plots instead of ribbons in the online system, in order to not clutter the score. In future research, we plan to set up an experiment in order to test if musicologists prefer the ribbon or curve representation.

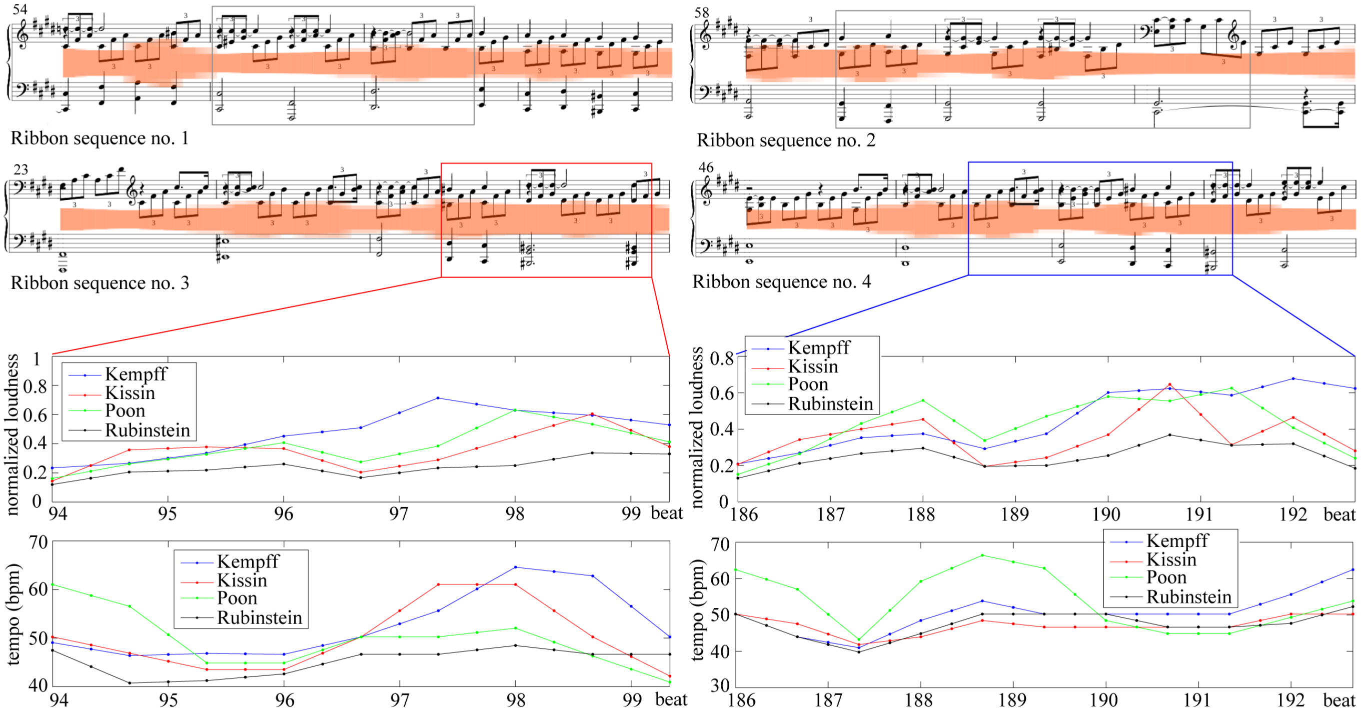
IV. Conclusions

In this paper, a web-based Interactive Multi-modal Music Analysis system (IMMA) that facilitates the fusion of audio and score based musical analysis is developed. The system performs audio-to-score alignment using a DTW-based algorithm. In a first use case, score- and audio based tension analysis modules were implemented. IMMA allows the user to visualize various aspects of tonal tension from score, synced with timbral tension features from an audio performance. We used the visualization and statistical analysis tools offered by IMMA to show the relationship between tonal and timbral tension. IMMA also includes a clustering algorithm that allows

[6] https://github.com/mechanicalscribe/vexflow-musicxml

[7] http://vexflow.com

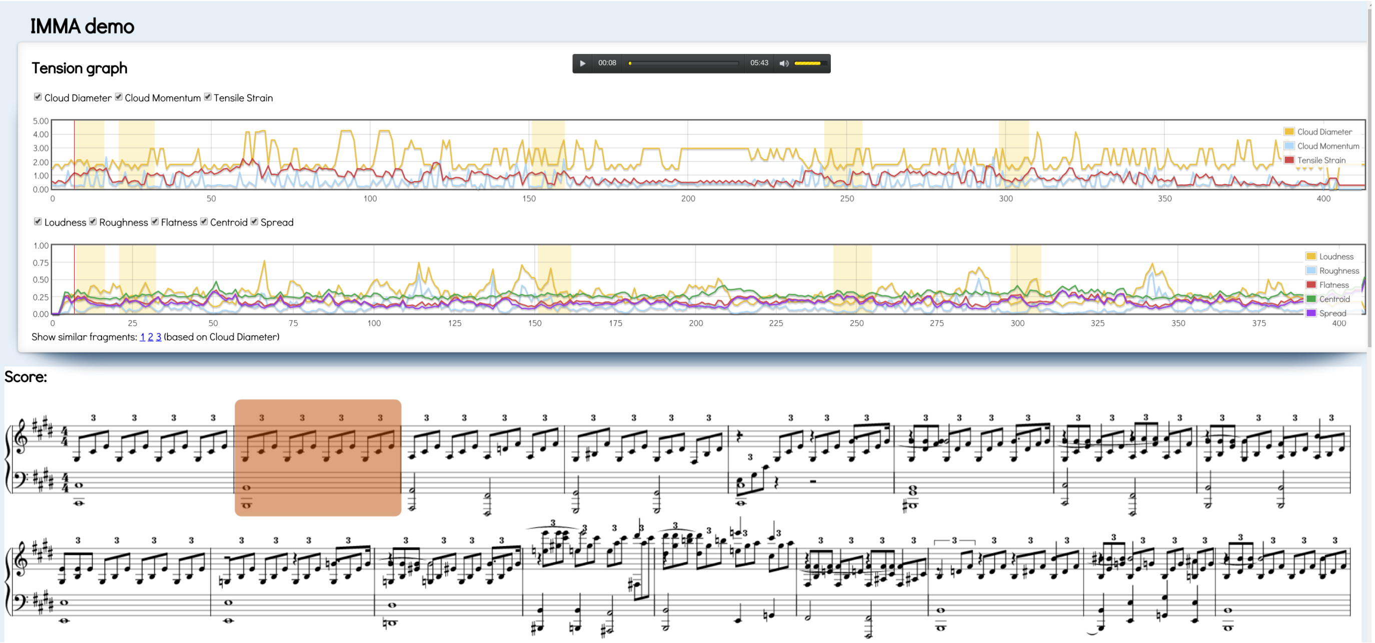
[8] http://flotcharts.org



(a) Loudness and tempo for the ribbon sequence no. 3

(b) Loudness and tempo for the ribbon sequence no. 4

Fig. 6: Loudness and tempo changes in the four performances of Beethoven's Sonata No. 14 in C♯ minor for the identified tri-gram sequences of cloud diameter tension ribbons.


IMMA demo

Tension graph

☑ Cloud Diameter ☑ Cloud Momentum ☑ Tensile Strain

☑ Loudness ☑ Roughness ☑ Flatness ☑ Centroid ☑ Spread

Show similar fragments: 1 2 3 (based on Cloud Diameter)

Score:

Fig. 7: The IMMA website.

us to compare segments of musical performances based on similar patterns in tension curves.

IMMA is implemented as an interactive web application that researchers in musicology and music cognition can use for their analyses. Many parameters can be customized by the researchers, depending on the purpose of their study, including the manner in which they describe the similarity between tension ribbon sequences.

It is widely acknowledged that cross-disciplinary collaboration is the key to the success of MIR research. However, such collaboration "challenges MIR to find a balance between features that are powerful but also make sense to collaborators who may not be experts in machine learning or audio signal processing" [26]. IMMA aims to provide a platform for researchers across different disciplines to systematically connect score and audio features to important semantic concepts in musicology and music cognition.

We will continue to improve IMMA by creating novel and state-of-the-art accessible interface features that allow the user to intuitively annotate music and provide feedback. In this manner, IMMA can be used as a tool for collecting semantic data from both experts and general users. Experiments will also be conducted to evaluate and further improve the usability of the system. An active learning developed by one of the authors [19] will be incorporated into IMMA in order increase the accuracy of audio-to-score alignment. The module-based back-end is built in such a way that it can easily be expanded with new modules and features that extend beyond tension analysis. Ultimately, we will work with researchers in musicology and music cognition who can use IMMA to explore new ways and directions in semantic music analysis.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 658914.

References

[1] M. Schedl, N. Orio, C. Liem, and G. Peeters, "A professionally annotated and enriched multimodal data set on popular music," in *Proceedings of the 4th ACM Multimedia Systems Conference*. ACM, 2013, pp. 78–83.

[2] J. Wang, H. Deng, Q. Yan, and J. Wang, "A collaborative model of low-level and high-level descriptors for semantics-based music information retrieval," in *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT'08. IEEE/WIC/ACM International Conference on*, vol. 1. IEEE, 2008, pp. 532–535.

[3] P. Saari and T. Eerola, "Semantic computing of moods based on tags in social media of music," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2548–2560, 2014.

[4] C. Cannam, C. Landone, M. B. Sandler, and J. P. Bello, "The sonic visualiser: A visualisation platform for semantic descriptors from musical signals." in *ISMIR*, 2006, pp. 324–327.

[5] C. L. Krumhansl, "A perceptual analysis of Mozart's piano sonata K. 282: Segmentation, tension, and musical ideas," *Music perception*, pp. 401–432, 1996.

[6] F. Lerdahl, "Calculating tonal tension," *Music Perception*, pp. 319–363, 1996.

[7] E. Bigand, R. Parncutt, and F. Lerdahl, "Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training," *Perception & Psychophysics*, vol. 58, no. 1, pp. 125–141, 1996.

[8] M. M. Farbood, "A parametric, temporal model of musical tension," *Music Perception*, vol. 29, no. 4, pp. 387–428, 2012.

[9] C. L. Krumhansl, "An exploratory study of musical emotions and psychophysiology." *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 51, no. 4, p. 336, 1997.

[10] W. E. Fredrickson, "Perception of tension in music: Musicians versus nonmusicians," *Journal of music Therapy*, vol. 37, no. 1, pp. 40–50, 2000.

[11] M. Lehne and S. Koelsch, "Tension-resolution patterns as a key element of aesthetic experience: psychological principles and underlying brain mechanisms," *Art, Aesthetics, and the Brain*, 2014.

[12] C. Palmer, "Anatomy of a performance: Sources of musical expression," *Music Perception: An Interdisciplinary Journal*, vol. 13, no. 3, pp. 433–453, 1996.

[13] M. M. Farbood and K. Price, "Timbral features contributing to perceived auditory and musical tension," in *Proceedings of the 13th International Conference on Music Perception and Cognition. Seoul, Korea*, 2014.

[14] D. Herremans and E. Chew, "Tension ribbons: Quantifying and visualising tonal tension," in *Second International Conference on Technologies for Music Notation and Representation (TENOR)*, Cambridge, UK, May 2016.

[15] J. Carabias-Orti, F. Rodriguez-Serrano, P. Vera-Candeas, N. Ruiz-Reyes, and F. Canadas-Quesada, "An audio to score alignment framework using spectral factorization and dynamic time warping," in *16th International Society for Music Information Retrieval Conference*, 2015.

[16] P. Cuvillier and A. Cont, "Coherent time modeling of semi-markov models with application to real-time audio-to-score alignment," in *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*. IEEE, 2014, pp. 1–6.

[17] C. Joder and B. Schuller, "Off-line refinement of audio-to-score alignment by observation template adaptation," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 206–210.

[18] R. J. Turetsky and D. P. Ellis, "Ground-truth transcriptions of real music from force-aligned midi syntheses," *ISMIR 2003*, pp. 135–141, 2003.

[19] C.-H. Chuan, "An active learning approach to audio-to-score alignment using dynamic time warping," in *Proceedings of the The 15th IEEE International Conference on Machine Learning and Applications*, 2016.

[20] E. Chew, *Mathematical and Computational Modeling of Tonality*. Springer, 2014.

[21] ——, "The spiral array: An algorithm for determining key boundaries," in *Proceedings of the Second International Conference on Music and Artificial Intelligence*. Springer-Verlag, 2002, pp. 18–31.

[22] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects*, 2007, pp. 237–244.

[23] R. T. Dean and W. T. Dunsmuir, "Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models," *Behavior research methods*, pp. 1–20, 2015.

[24] M. Good, "Musicxml for notation and analysis," *The virtual score: representation, retrieval, restoration*, vol. 12, pp. 113–124, 2001.

[25] B. Peiris, *Instant JQuery Flot Visual Data Analysis*. Packt Publishing Ltd, 2013.

[26] A. Honingh, J. A. Burgoyne, P. van Kranenburg, A. Volk *et al.*, "Strengthening interdisciplinarity in mir: Four examples of using mir tools for musicology," *ILLC Publications, Prepublication Series*, 2014.