# IMMA-Emo: A Multimodal Interface for Visualising Score- and Audio-synchronised Emotion Annotations

**Dorien Herremans***
Singapore University of Technology and Design, Information Systems and Design Pillar
Singapore
dorien.herremans@gmail.com

**Simin Yang**
Queen Mary University of London, Centre for Digital Music
London, UK
simin.yang@qmul.ac.uk

**Ching-Hua Chuan**
University of North Florida
Jacksonville, USA
c.chuan@unf.edu

**Mathieu Barthet**
Queen Mary University of London, Centre for Digital Music
London, UK
m.barthet@qmul.ac.uk

**Elaine Chew**
Queen Mary University of London, Centre for Digital Music
London, UK
elaine.chew@qmul.ac.uk

## ABSTRACT

Emotional response to music is often represented on a two-dimensional arousal-valence space without reference to score information that may provide critical cues to explain the observed data. To bridge this gap, we present IMMA-Emo, an integrated software system for visualising emotion data aligned with music audio and score, so as to provide an intuitive way to interactively visualise and analyse music emotion data. The visual interface also allows for the comparison of multiple emotion time series. The IMMA-Emo system builds on the online interactive Multi-modal Music Analysis (IMMA) system. Two examples demonstrating the capabilities of the IMMA-Emo system are drawn from an experiment set up to collect arousal-valence ratings based on participants' perceived emotions during a live performance. Direct observation of corresponding score parts and aural input from the recording allow explanatory factors to be identified for the ratings and changes in the ratings.

*D. Herremans performed part of this research at Queen Mary University of London.


*AM '17, August 23–26, 2017, London, United Kingdom*

ACM ISBN 978-1-4503-5373-1/17/08…$15.00
https://doi.org/10.1145/3123514.3123545

## CCS CONCEPTS

• **Information systems → Multimedia and multimodal retrieval**; • **Human-centered computing → Information visualization**;

## KEYWORDS

music and emotion, arousal/valence, emotion visualisation, computational musicology, multimodal user interface

**ACM Reference format:**
Dorien Herremans, Simin Yang, Ching-Hua Chuan, Mathieu Barthet, and Elaine Chew. 2017. IMMA-Emo: A Multimodal Interface for Visualising Score- and Audio-synchronised Emotion Annotations. In *Proceedings of AM '17, London, United Kingdom, August 23–26, 2017,* 8 pages.
https://doi.org/10.1145/3123514.3123545

## 1 INTRODUCTION

Driven by a recent rise in technologies such as automatic genre classification and music recommendation, there has been an increased interest in music emotion models [4, 46]. The availability of large volumes of digital music has made it possible to develop complex computational models that aim to capture the relationship between music and listeners' emotional responses [5, 12, 27, 41, 44, 45]. Integrating these models of emotion can greatly contribute to the quality and usability of recommender systems.

Visualisation can be an important tool to understand the relationship between music and emotion. There exist several software tools for semantic music analysis that attempt to associate human-understandable meanings to music signals [9]; some have focussed on score- *and* audio-based synchronised visualisation [2, 11, 28, 42], but none yet integrate listeners' emotion data in their visualisations. To address this issue, we have developed IMMA-Emo, an extension of the IMMA system, a web-based system for music analysis that synchronises both audio playback and score tracking together with the visualisation of music analysis data [20]. This extension augments the existing IMMA system with real-time arousal-valence [33] visualisation. A demo of the new module, IMMA-Emo, is available online[1]. We believe that an interface with visualisations of listeners' emotional responses temporally synchronised with score information and the audio will benefit psycho-musicological analysis of music and emotion information and be intuitive for users. The insights obtained from an audiovisual analysis of the data can then serve as input to the design of explanatory models for music and emotion.

Section 2 gives an overview of related research concerning music emotion modelling and visualisation. Section 3 describes the technical details of the IMMA system and how it was expanded to visualise emotion data. To demonstrate the new system and its use in score- and audio-informed music emotion analysis, a study was set up to collect perceived emotion [17, 23] ratings from 15 listeners during a live performance. The case study results are visualised in the demo website and discussed in Section 4. Section 5 concludes the paper with a summary of our main contributions and a description of future possible directions.

## 2 RELATED WORK

This section presents a brief overview of research on music and emotion modelling including studies addressing temporal aspects. We also discuss several existing visualisation systems for music analysis.

### Modelling emotion in music

Music has the ability to induce emotions in listeners (known as felt emotions) and/or convey emotions (known as perceived emotions) [17, 23]. Research on music and emotion modelling has grown over the past two decades. Drawing on from results in psychology, representations of emotion in music emotion recognition (MER) models have to date mainly followed a categorical (discrete) or dimensional (continuous) approach [4]. In categorical representations emotions are viewed as discrete categories (e.g. happy or sad [13, 32]). A disadvantage of categorical approaches is that they constrain emotions to a set of predefined discrete families or landmarks [30], which also makes it more difficult to cross-reference studies for comparison [24]. In dimensional representations, emotions can be characterised by degrees or levels along scales. As the dimensional emotion representation has been shown to provide greater accuracies for music emotion recognition [4] it is often preferred over the categorical representation in MER studies (see e.g. [5, 6, 22, 26, 38, 40, 48]).

Amongst dimensional models [33, 35, 39] the most widely used in emotion research is arguably Rusell [33]'s circumplex model of affect. This model underlies a two-dimensional representation of emotions with valence (degree of pleasantness) and arousal (degree of energy) as scales. This representation is often referred to as the arousal-valence (AV) space. We have opted to use such AV representation of emotion in the IMMA-Emo interface, not only because of its robustness and intuitiveness [4] but also due to its ability to visualise the dynamics of emotion variation over time. Validating emotion models suitable for music is not an easy task. It requires hand annotating of large musical datasets and verification by musicologists. Recently, more complex models of emotion are developed using psycho-physiological correlates of emotions such as electromyogram (EMG), electrocardiogram (ECG), skin conductance linked to electro-dermal activity (EDA), respiration changes and skin temperature [25, 29, 34, 43].

In this paper, we develop a platform for studying emotional response of listeners through the visualisation of ratings through a user-friendly interface that allows us to readily apprehend the temporal aspects of music.

### Importance of time in music emotion models

As music unfolds in time the temporal aspect of music is inextricably tied to any musical experience. Studies suggest that listeners' emotional response to music is tightly linked to the temporal evolution of music features [21, 38]. However, to date, the time-varying nature of music and its impact on emotional response has only been scarcely addressed in MER [8]. Most MER systems focus on assigning one emotion label to a whole piece or long-term excerpts without taking into account the time-dependent nature of emotional response to music. As a result, listeners' emotion ratings are usually collected as single ratings for each stimulus (e.g. the NTUMIR dataset [47]), which is counter to the fact that perceived emotion changes over time as music unfolds [36]. In other works by Schmidt et al. [37], and Panda and Paiva [31], a musical piece is first segmented before collecting emotion ratings for each segment. This accounts partially for time variability in emotion ratings; and, not surprisingly, has been shown to result in markedly better models and results.

In order to better understand, and visualise the time-based effects of music on perceived emotions, the interface proposed in this research offers a representation that visualises the score in synchrony with arousal-valence data and audio playback. We hope that such audiovisual user interface can assist music emotion and musicologist researchers to uncover relationships between (temporal) music score- and

[1]http://dorienherremans.com/imma

audio-based features and emotional response, with implications for the design of MER models.

### Music analysis and visualisation

Music analysis is an integral part of music information retrieval and musicological studies. In recent decades, the availability of large digital music datasets and new technologies has spurred the development of computer systems for automated and semi-automated analysis of music. This has led to a recent trend in systematic or computational musicology, where musicological findings are based on ever larger amounts of data using computational methods [10]. Such aspirations and advances in big data research have provided grounds on which to develop large-scale analysis and visualisation tools to assist in the interpretation and verification of musicological findings, such as that proposed by the AHRC Digital Music Lab project [1]. While many software systems allow for semantic music analysis from audio, none yet integrate arousal-valence information comprehensively in their workflows and visualisations. Sonic Visualiser (SV) [9], for instance, lets users annotate and visualise acoustic features in audio recordings. Displaying or analysing the musical score in SV is only supported in a “piano roll" type of representation used in contemporary music production but not suiting musicological analysis; Also, being a desktop application SV, does not have the advantages that web-based platforms have in terms of accessibility and connectivity (linked data).

In score-based systems, we find MuseScore [7], a powerful score editor that allows the user to annotate a score; MuseScore allows for synchronisation of score to video, but does not yet support automatic music analysis or performance synchronisation. Automatic systems for audio- and score-synchronisation include [2, 11, 14, 28, 42].

Barthet and Dixon [3] argues for the importance of an integrated multimodal music representation for musicological research. Their conclusions were based on observations that musicologists frequently use different sources of musical content (e.g. YouTube for the music, paper scores) requiring them to constantly switch between media and technologies. The IMMA system developed by Herremans and Chuan [20] is designed to bridge the gap between audio- and score-based music analysis. Its online interface allows for the synchronised visualisation of different types of features. The extension module developed in the current research integrates a function for displaying emotion data.

## 3 THE IMMA-EMO SYSTEM

### The IMMA System

IMMA is a web-based interactive multimodal interface for visualising music analysis data. It was developed to bridge the gap between audio- and score-based music analysis [20]. The initial module of IMMA focused on visualising tension extracted from both audio- [15] and score-based [19] characteristics. In this work, we develop a new IMMA module for arousal-valence visualisation.

The IMMA interface is built in Javascript with jQuery, using a modular design that allows for easy extension. The score is loaded from a musicXML file, an open format built for easy interchange of scores [18], through the vexFlow API[2]. The parsing is handled by the VexFlow MusicXML plugin[3]. VexFlow renders the score on an HTML5 canvas with scalable vector graphics support using Javascript and jQuery. The score rendering in the current version of VexFlow MusicXML is still work in progress; features such as slurs and visual separation of notes and lines are not yet fully captured from the MusicXML file. Future versions may have improved score rendering capabilities. Audio features are displayed in interactive plots using the Flot Charts API[4]. Herremans and Chuan [20] describe the dynamic time-warping approach implemented in IMMA to automatically synchronise a performance with its score. This allows for the synchronisation of audio playback with score following and music analysis data.

### The IMMA-Emo module

The emotion visualisation module for IMMA consists of two panels. The first consists of two graphs, each visualising one dimension of the emotion annotations (or ratings) over time, see Figure 1. Each annotator is plotted using a different colour in the graph. A crosshair vertical line indicator marks the current playhead position in this graph during audio playback.

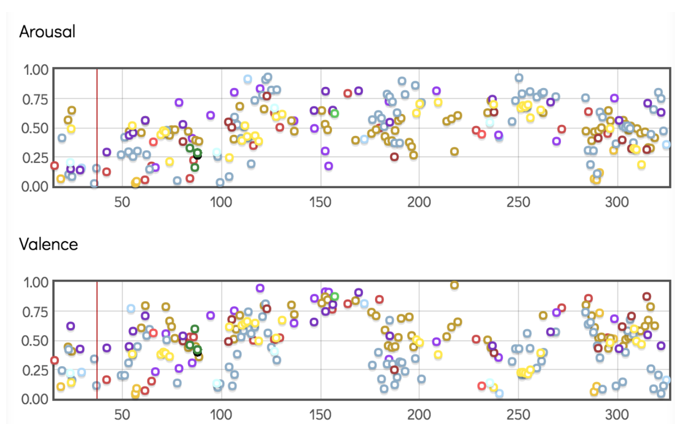


**Figure 1: Plot of arousal-valence data over time, with crosshair synced to playback and score following.**

[2] http://vexflow.com

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

[4] http://flotcharts.org

Secondly, a two-dimensional arousal-valence graph displays the current annotation for each annotator as a circle (Figure 2). During playback, circles either remain at their last assignments when listeners do not change their ratings, or move to new positions if changes occur at playhead time.

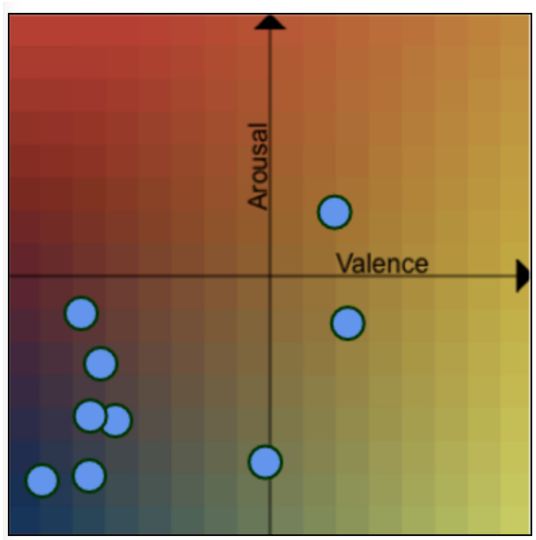


**Figure 2: Two dimensional arousal-valence graph with annotations of all participants.**

In the next section, the IMMA-Emo module is demonstrated through a case study.

## 4 CASE STUDY: BABAJANIAN TRIO

The IMMA-Emo interface is demonstrated through annotation data of perceived emotion collected during a live concert. The annotations together with the recorded audio and the score of the performed piece in MusicXML format serve as input to the IMMA-Emo system for audiovisual analysis of the emotional response to the performed music.

### Study Design

A study was set up to collect data for exploratory analysis of listeners' perceived emotions over the course of a live performance. The study was held as part of the Art Inside Music concert[5] (London Inside Out Festival) on 22 October 2015 in the Media and Arts Technology[6] Performance Lab at Queen Mary University of London. The music consisted of the three movements of the *Piano Trio in F# minor by Arno Babajanian (1921 - 1983)*. The three movements were of widely disparate characters: the first is marked Largo – Allegro expressivo – Maestoso, meaning it is largely in a slow and noble tempo with a faster middle part; the second is marked Andante, meaning it is performed at a walking pace; and, the third is marked Allegro vivace, which is lively and rapid. The entire trio is about 23 minutes in length, and was performed twice by Hilary Sturt (violin), Ian Pressland (cello), and Elaine Chew (piano), all expert musicians. Consenting audience members were invited to report the emotions they perceived in the music whenever they felt a change using the smartphone-friendly Mood Rater web application[7]. Participants were introduced to the two-dimensional arousal/valence interface displayed in the app. No specific indications were given to participants on how frequently they should rate. We assume that given emotion annotations last until a change is reported. After the performance, the audience members were asked to complete a short questionnaire in order to collect demographics information and help us improve the app. The annotated data used in the examples to follow are from the first movement of the first performance.

### Results

Amongst a total of about 50 audience members, 15 participants actively used the app to report their perceived emotion ratings during the concert and 13 completed the questionnaire at the end. The age of the participants ranged from 23 to 36 years (age 20-24: 2; age 25-29: 8; age 30-35: 2; age 36-40: 1), with a well balanced male/female ratio (male: 6, female: 7). The resulting AV annotations, the musical score and audio of the performance are all visualised in the IMMA-Emo interface which is available online[8]. IMMA-Emo has an automatic function to synchronise a recorded performance with the corresponding score. However, manual annotation still outperforms automated audio-score alignment methods especially for late Romantic repertoire with fast and extreme rubato and tempo changes as in the case of the Babajanian Trio. For the best results, we manually annotated the data on a bar-level, using Sonic Visualiser to minimize synchronisation errors. Two excerpts are discussed below.

### Example 1: Constant Emotion

Figure 3 displays a screenshot of our interface 36 seconds into the recording. As shown in the AV plane and graphs, the majority of the participants' perceived emotions fell in the the lower left quadrant, i.e. most participants reported low perceived valence and arousal. At this point, the performers have reached bar 4 in the introductory *Largo* section of the first movement, as highlighted in the score. The tempo is

[5]https://sites.google.com/site/c4dmconcerts1516/home/classical/artinsidemusic
[6]http://www.mat.qmul.ac.uk/

[7]http://bit.ly/moodxp2
Originally developed as part of the the "Mood Conductor" framework for participatory music performance [16], this web application has been here re-purposed and re-branded as "Mood Rater" for live emotion data collection.
[8]Select arousal/valence demo at: http://dorienherremans.com/imma

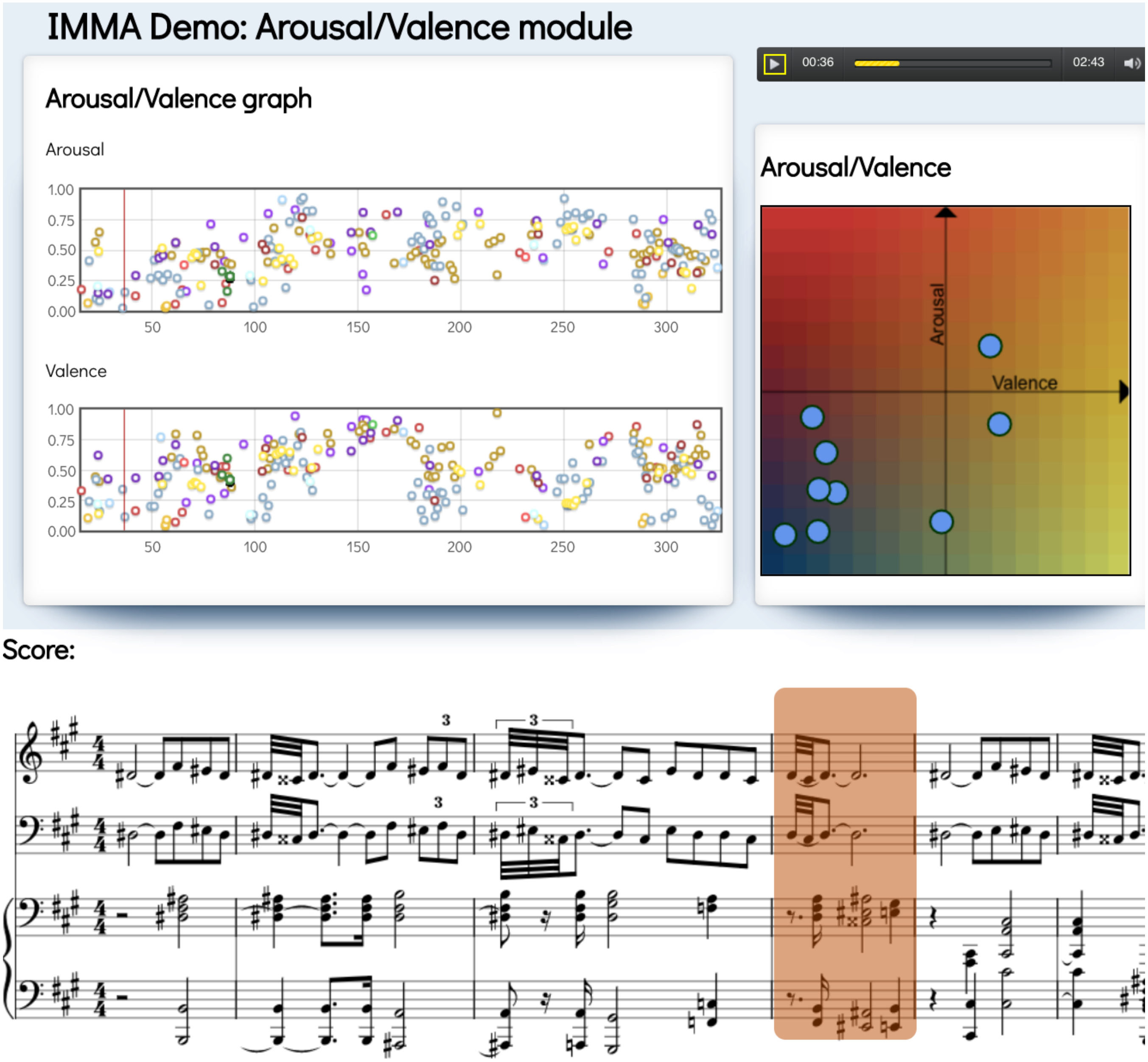


**Figure 3: Example 1 - IMMA-Emo screenshot in the midst of a passage leaning towards low arousal and low valence (36 seconds into the recording.** ***Note: The top left Arousal/Valence graph displays the arousal and valence ratings over time. Current playback time is indicated by a vertical playhead. The top right graph displays the AV ratings from all listeners in a 2D space. In this 2D space, ratings correspond to listeners' last ratings when unchanged, or to new ratings that occur at playhead time.***

slow, the dynamics low, and the music is in the lower registers on all instruments. The score was input accurately into MuseScore and made as close as possible to the original; errors in note alignment between the parts, missing accidentals when accidentals are repeated in the same staff or across parts, and missing slurs in the score panel are introduced due to the score rendering limitations of VexFlow or the VexFlow XML parser.

### Example 2: Moments of Change

An example that shows how arousal and valence ratings varied as music unfolded can be observed at around 100 – 120 seconds into the recording. The screenshot in Figure 4 shows

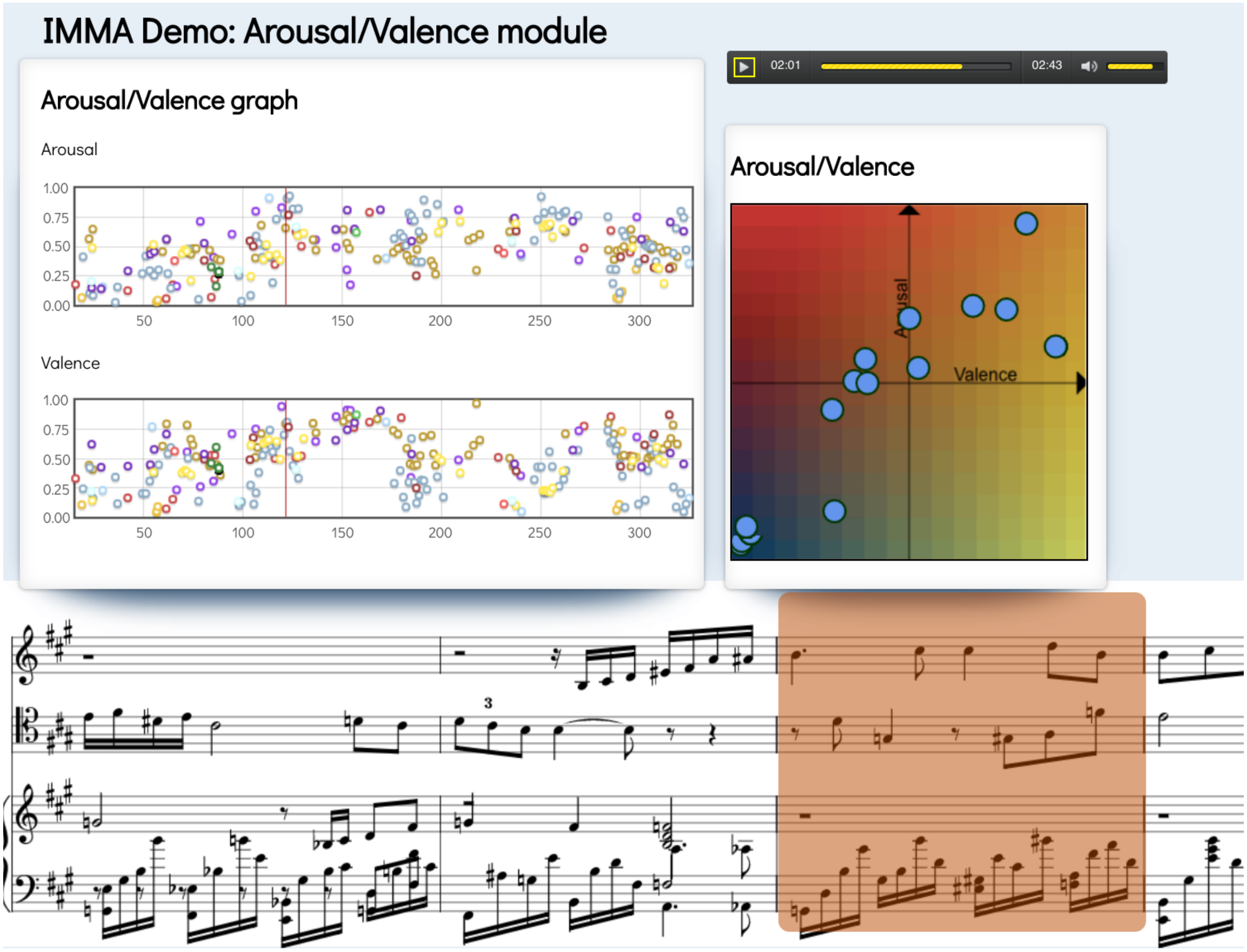


**Figure 4: Example 2 - IMMA-Emo screenshot at a point of perceived emotion change from low to high arousal/valence (121 seconds into the recording).** ***Note: The top left Arousal/Valence graph displays the arousal and valence ratings over time. Current playback time is indicated by a vertical playhead. The top right graph displays the AV ratings from all listeners in a 2D space. In the 2D space, ratings correspond to listeners' last ratings when unchanged, or to new ratings that occur at playhead time. More ratings appear in the AV space in this figure compared to Figure 3, as not all participants had made initial ratings in Example 1.***

the moment when many of the valence as well as arousal listener ratings suddenly increased, indicating that the participants perceived a more positive mood and higher energy in the music. The passage shown is from the fluid *Allegro espressivo* (middle) section of the first movement. This section begins with the cello introducing the lyric second theme over the steady stream of semi-quavers in the piano accompaniment, and the moment shown is immediately after the violin takes over the theme from the cello, rising out of the cello registers into the higher violin range. IMMA-Emo can help forging hypothes(e)s for the observed change of perceived emotion using audio, annotation and score information in a combined way. In this instance, the shift to the top right quadrant may come from the fact that listeners pay attention to, and react to, the introduction of the violin voice into the ensemble texture.

## 5 CONCLUSIONS

A novel module for the web-based multimodal music analysis system IMMA was created for visualising emotion annotations. This module, called IMMA-Emo, allows for a score- and audio-synchronised visualisation of emotion ratings in the two-dimensional arousal-valence space. A second panel displays plots of arousal and of valence data over time, with a crosshair synchronising with the audio playback and score following.

The new interface was tested with data from a concert study. Perceived emotion annotations were gathered during a live performance of the Piano Trio in F# minor by Arno Babajanian. The results were displayed and analysed in IMMA. The new interface allows for easy temporal and audiovisual analysis of the data. We believe that such tool has the potential to support data-driven investigations of musicologists, music information researchers, music analysts, and researchers in music perception and cognition.

The modular design of the IMMA platform readily allows for the concurrent visualisation of other music features. Future extensions to the interface will include designs that incorporate tension, loudness, and tempo data. The planned integration of an upload function will also allow IMMA to grow into an online repository for music analysis data. Forthcoming studies can use the insights gleaned from the interface to create analytical models that connect music and audio features to perceived emotions. Detailed case studies can also be introduced to formally validate the improvements the interface brings to the music emotion research workflow.

## 6 ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 658914. Simin Yang acknowledges the support of the China Scholarship Council. Mathieu Barthet acknowledges support from the EU H2020 Audio Commons grant (688382).

## REFERENCES

[1] Samer Abdallah, Emmanouil Benetos, Nicolas Gold, Steven Hargreaves, Tillman Weyde, and Daniel Wolff. 2017. The Digital Music Lab: A Big Data Infrastructure for Digital Musicology. *J. Comput. Cult. Herit.* 10, 1, Article 2 (Jan. 2017), 2:1–2:21 pages. https://doi.org/10.1145/2983918

[2] Andreas Arzt, Werner Goebl, and Gerhard Widmer. 2015. Flexible Score Following: The Piano Music Companion and Beyond. In *Proc. of the Third Vienna Talk on Music Acoustics.* University of Music and Performing Arts Vienna, 220–223.

[3] Mathieu Barthet and Simon Dixon. 2011. Ethnographic Observations of Musicologists at the British Library: Implications for Music Information Retrieval.. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR).* 353–358.

[4] Mathieu Barthet, György Fazekas, and Mark Sandler. 2012. Multidisciplinary perspectives on music emotion recognition: Implications for content and context-based models. *Proc. CMMR* (2012), 492–507. http://cmmr2012.eecs.qmul.ac.uk/sites/cmmr2012.eecs.qmul.ac.uk/files/pdf/papers/cmmr2012_submission_101.pdf

[5] M. Barthet, D. Marston, C. Baume, Fazekas G., and M. Sandler. 2013. Design and Evaluation of Semantic Mood Models for Music Recommendation Using Editorial Tags. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR).* 421–426.

[6] C. Baume, G. Fazekas, M. Barthet, D. Marston, and M. Sandler. 2014. Selection of audio features for music emotion recognition using production music. In *Proc. of Audio Eng. Society 53rd Conference.*

[7] T Bonte. 2009. MuseScore: Open source music notation and composition software. Technical report, Free and Open source Software Developers' European Meeting, 2009. http://www. slideshare. net/thomasbonte/musescore-at-fosdem-2009.

[8] Marcelo Caetano, Athanasios Mouchtaris, and Frans Wiering. 2012. The role of time in music emotion recognition: Modeling musical emotions from time-varying music features. In *Int. Symposium on Computer Music Modeling and Retrieval.* LNCS Springer, 171–196.

[9] Chris Cannam, Christian Landone, Mark B Sandler, and Juan Pablo Bello. 2006. The Sonic Visualiser: A Visualisation Platform for Semantic Descriptors from Musical Signals.. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR).* 324–327.

[10] Eric Clarke and Nicholas Cook. 2004. *Empirical musicology: Aims, methods, prospects.* Oxford University Press.

[11] Antonello D'Aguanno and Giancarlo Vercellesi. 2007. Automatic synchronisation between audio and score musical description layers. In *Proc. Int. Conference on Semantic and Digital Media Technologies.* Springer, 200–210.

[12] Tuomas Eerola and Jonna K Vuoskoski. 2011. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music* 39, 1 (2011), 18–49.

[13] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.

[14] Alice Eldridge, Ed Hughes, and Chris Kiefer. 2016. Designing dynamic networked scores to enhance the experience of ensemble music making. In *Proc. Second International Conference on Technologies for Music Notation and Representation.* 193–199.

[15] Morwaread Mary Farbood and Khen Price. 2014. Timbral features contributing to perceived auditory and musical tension. In *Proc. of Int. Conference on Music Perception and Cognition.*

[16] Gyorgy Fazekas, Mathieu Barthet, and Mark B Sandler. 2013. The Mood Conductor System: Audience and Performer Interaction using Mobile Technology and Emotion Cues. In *Proc. Int. Symposium on Computer Music Multidisciplinary Research (CMMR).* 15–18.

[17] Alf Gabrielsson. 2001. Emotion perceived and emotion felt: Same or different? *Musicae Scientiae* 5, 1_suppl (2001), 123–147.

[18] Michael Good. 2001. MusicXML for notation and analysis. *The virtual score: representation, retrieval, restoration* 12 (2001), 113–124.

[19] Dorien Herremans and Elaine Chew. 2016. Tension ribbons: Quantifying and visualising tonal tension. In *Proc. Int. Conf. on Technologies for Music Notation and Representation (TENOR).* Cambridge, UK.

[20] Dorien Herremans and Ching-Hua Chuan. 2017. A multi-modal platform for semantic music analysis: visualizing audio- and score-based tension. In *Proc. IEEE Int. Conference on Semantic Computing (ICSC).* San Diego, US.

[21] Rumi Hiraga and Noriyuki Matsuda. 2004. Graphical expression of the mood of music. In *IEEE Int. Conf. on Multimedia and Expo (ICME)*, Vol. 3. IEEE, 2035–2038.

[22] Arefin Huq, Juan Pablo Bello, and Robert Rowe. 2010. Automated music emotion recognition: A systematic evaluation. *Journal of New Music Research* 39, 3 (2010), 227–244.

[23] Patrik N Juslin and Petri Laukka. 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research* 33, 3 (2004), 217–238.

[24] Patrik N Juslin and Daniel Västfjäll. 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences* 31, 05 (2008), 559–575.

[25] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence* 30, 12 (2008), 2067–2083.

[26] Youngmoo E Kim, Erik M Schmidt, and Lloyd Emelle. 2008. Moodswings: A collaborative game for music mood label collection.. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*, Vol. 2008. 231–236.

[27] Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. 2010. Music emotion recognition: A state of the art review. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*. Citeseer, 255–266.

[28] Frank Kurth, Meinard Müller, Christian Fremerey, Yoon-ha Chang, and Michael Clausën. 2007. Automated Synchronization of Scanned Sheet Music with Audio Recordings. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*. 261–266.

[29] Richard A McFarland. 1985. Relationship of skin temperature changes to the emotions accompanying music. *Applied Psychophysiology and Biofeedback* 10, 3 (1985), 255–267.

[30] Marcello Mortillaro, Ben Meuleman, and Klaus R Scherer. 2012. Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions (IJSE)* 3, 1 (2012), 18–32.

[31] Renato Panda and Rui Pedro Paiva. 2011. Using support vector machines for automatic mood tracking in audio music. In *Audio Engineering Society Convention 130*. Audio Engineering Society.

[32] Jaak Panksepp. 2004. *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press.

[33] J Rusell. 1980. A circumplex model of affect. *Personality and Social Psychology* (1980), 1161–1178.

[34] Valorie N Salimpoor, Mitchel Benovoy, Kevin Larcher, Alain Dagher, and Robert J Zatorre. 2011. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature neuroscience* 14, 2 (2011), 257–262.

[35] Ulrich Schimmack and Alexander Grob. 2000. Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality* 14, 4 (2000), 325–345.

[36] Erik M Schmidt and Youngmoo E Kim. 2010. Prediction of Time-varying Musical Mood Distributions from Audio. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*. 465–470.

[37] Erik M Schmidt, Douglas Turnbull, and Youngmoo E Kim. 2010. Feature selection for content-based, time-varying musical emotion regression. In *Proc. Int. Conference on Multimedia Information Retrieval*. ACM, 267–274.

[38] Emery Schubert. 2004. Modeling perceived emotion with continuous musical features. *Music Perception* 21, 4 (2004), 561–585.

[39] Emery Schubert. 2007. The influence of emotion, locus of emotion and familiarity upon preference in music. *Psychology of Music* 35, 3 (2007), 499–515.

[40] Björn W Schuller, Felix Weninger, and Johannes Dorfner. 2011. Multi-Modal Non-Prototypical Music Mood Analysis in Continuous Space: Reliability and Performances.. In *Proc. of Int. Society for Music Information Retrieval Conference (ISMIR)*. 759–764.

[41] Yading Song, Simon Dixon, Marcus T Pearce, and Andrea R Halpern. 2016. Perceived and Induced Emotion Responses to Popular Music. *Music Perception* 33, 4 (2016), 472–492.

[42] Verena Thomas, Christian Fremerey, Meinard Müller, and Michael Clausen. 2012. Linking Sheet Music and Audio—Challenges and New Approaches. In *Multimodal Music Processing*, M. Müller, M. Goto, and M. Schedl (Eds.). Dagstuhl Follow-Ups, Vol. 3. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 1–22.

[43] K. Trochidis, D. Sears, D.-L. Tran, and S. McAdams. 2012. Psychophysiological measures of emotional response to Romantic orchestral music and their musical and acoustic correlates. In *Proc. 9th Int. Symposium on Computer Music Modelling and Retrieval (CMMR)*.

[44] Konstantinos Trohidis, Grigorios Tsoumakas, George Kalliris, and Ioannis P Vlahavas. 2008. Multi-Label Classification of Music into Emotions.. In *ISMIR*, Vol. 8. 325–330.

[45] Tien-Lin Wu and Shyh-Kang Jeng. 2008. Probabilistic estimation of a novel music emotion model. *Advances in Multimedia Modeling* (2008), 487–497.

[46] Yi-Hsuan Yang and Homer H Chen. 2011. *Music emotion recognition*. CRC Press.

[47] Yi-Hsuan Yang and Homer H Chen. 2011. Prediction of the distribution of perceived music emotions using discrete samples. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 7 (2011), 2184–2196.

[48] Yi-Hsuan Yang, Chia-Chu Liu, and Homer H Chen. 2006. Music emotion classification: A fuzzy approach. In *Proceedings of the 14th ACM international conference on Multimedia*. ACM, 81–84.