MIST: "You Play, I'll Draw"

Juliana Shihadeh and Margareta Ackerman
Department of Computer Science and Engineering
Santa Clara University, Santa Clara, California
{jshihadeh,mackerman}@scu.edu

Abstract

What if creative machines could show the story told by a live orchestra? In this paper, we propose a vision for a co-creative orchestra performance partner that enlivens this classical art form and has the potential to increase audience engagement and reach broader audiences. We present MIST, an early version of our vision, which creates real-time visualizations as live music is performed, and showcase it in a violin performance.

Introduction

Collaboration between humans and machine opens the possibility of creative entertainment that neither humans nor machines could create alone. We propose that human-machine creative collaboration can be used to enliven classical art performances. Orchestral concert attendance is on the decline, from 11.2% in 2002 to 9.1% in 2008 (Williams and Keen 2008). Through a perhaps unlikely juxtaposition of the old and the new, Computational Creativity may offer an opportunity to enliven engagement in classical performance.

Music is multidemensional, carrying structural features such as tempo, tone, pitch, range, and rhythm (Schaefer 2017). This provides ample information to extract and decrypt, giving the opportunity to translate the musical data into illustrations that unpack what music is telling through sound and bring forth a new dimension to the performance, allowing for deeper audience engagement.

In this paper, we present a vision for a novel form of collaboration between humans and machines, where the computer agent creates a visual narrative representation of music performed in real time - different every time the music is played. The machine partner would interpret the music, creating a visual story inspired by it, with the aim of deepening audience engagement though this visual medium.

One of the primary challenges with translating live sound into illustrations is real-time analysis of the music. There are many ways to visualize illustrations that represent the story told by an orchestra performance. Here we initiate this line of research, presenting an initial version of MIST where music is translated into sequences of symbols and emojis[1] to showcase real-time music extraction and illustration with smooth transitions between each music-to-illustration captured. We demonstrate the current version of the system through a collaboration with a human violin player.

In addition to presenting an early version of MIST, we discuss our vision for this line of work, proposing a partnership between an orchestra and a system illustrating the story conveyed through the live music (See Figure 1 for an illustration for the vision for this work).

Background

The idea of utilizing creative systems to enhance live performance of human artists has been applied in other manifestations in music and dance. For example, in live dance performance, ViFlow was used to create visuals in response to a dancer's movement, showcased behind or below the dancer during a performance (Brockhoeft et al. 2016). Other work related to integration of music and visuals includes (Chen et al. 2008), who created a system for the integration of user-provided music and photographs through an emotion-based approach.

Improvisation systems have also been applied to live performances. For example, the musical composition system Iamus (Ball 2012) built off the Melomics hardware (Diaz-Jerez 2011) has been used to generate music in real-time, played by an orchestra or human musicians. Robert Keller's improvisor has been used for musical human-machine improvisation trading (Putman and Keller 2015). Likewise, Andrew Brown's (Brown 2018) musical bot partnered with a human pianist to play a live improvisational duet and Weinberg et al's (Weinberg and Driscoll 2006) system Haile analyzes music in real-time and plays percussion instruments alongside the musicians.

MIST offers a new form of creative interaction between humans and machines, giving live musicians the opportunity to enrich their performances with improvisational visuals to complement the music, while giving audiences new dimensions through which to experience the performance.

[1] It is worth noting that due to their central role in youth culture, emojis have been incorporated into and inspiring art for several years https://news.yahoo.com/trend\-art\-inspired\-emojis\-emoji\-inspired\-art\-203634622.html

Proceedings of the 12th International Conference on Computational Creativity (ICCC '21)
ISBN: 978-989-54160-3-5

Figure 1: An illustration of the human-machine collaboration proposed in this work: While an orchestra plays, an animated story based on the live music is automatically created and rendered onto a screen. This offers a visual dimension to the orchestra experience and stands to increase audience engagement.

MIST

Our initial version of MIST is an interactive system that analyzes music in real-time and creates symbol-based moving illustrations using on information extracted from the music. We create two different visualization styles, one which consists of a sequence of emojis created on the fly in response to the music, and another which generates a sequence of colored stars. A demonstration of MIST interacting with a live violin player can be seen here: `https://youtu.be/sFeHJ6DBXcQ`.

Three Elements of a Live Music Illustration Systems

We propose three elements that a creative system that illustrates live music should include. These elements can enable a system to capture and represent the essence portrayed in the music.

1. Extract musical elements such as dynamics, pitch, beat, etc.
2. Listen and react to music played in real time
3. Convert the music to visual illustrations

The Emoji and Colored Stars prototypes demonstrate how features in music can be extracted to draw illustrations in real-time.

Emojis version:

As music is played, Emojis print out based on the audio's amplitude and pitch. We use Librosa's melspectrogram to determine amplitude[2] and CREPE[3] to extract frequency to determine octave[4]. To be able to extract the amplitude and pitch of the music in real-time, we use SoundDevice[5] to read audio in real-time, save to a wave file and then open it as if it was already pre-recorded. We divide amplitudes and octaves into ranges to map to an Emoji, first categorizing by amplitude and then the octave.

Softer amplitudes are mapped to emojis like flowers, leaves and hearts. Mid-range amplitudes focus on emojis that represented the "peak" of a story such as revolving or broken hearts. Higher-ranges of amplitude focus on a reflective, ponderous feeling as the aftereffect to the "peak" with clouds and umbrellas used to represent a rainy day where someone might be lost in thought or a waxing crescent moon to represent a new start, since high amplitude could also indicate an enthusiastic reflection of someone looking forward to the future.

Within amplitudes, we pick the final emoji based on the octave. We assign lower octaves emojis with a feeling of sadness or questioning of hope such as a flying balloon symbolising "leaving" or a seed representing "questioning hope" as one cannot really know how a seed's growth will turn out. Higher octaves range from showing emojis giving a feeling of deep sadness or excitement depending on the amplitude using emojis like clouds or bees.

Understanding the emotion music portrays requires insight into the complex relationship between musical elements. Naturally, illustrations can vary vastly in the way they tell the story of a piece of music and one mapping offers one such option. Figure 2 is of a segment of Vivaldi's Winter 1st Movement where the solo violinist is playing in a high octave with instruments playing softly in the background[6] (Lockey 2017).

Please note that rather than creating static images, MIST generates the sequences of emojis in syncrony with the music, and as such is best viewed in video form. We

[2]https://librosa.org/doc/latest/index.html
[3]https://pypi.org/project/crepe/
[4]https://pages.mtu.edu/ suits/notefreqs.html
[5]https://python-sounddevice.readthedocs.io/en/0.4.1/
[6]https://www.youtube.com/watch?v=JkP7slIc9aM

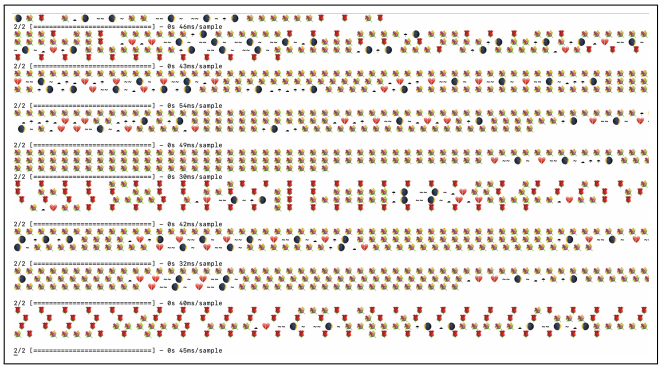

Figure 2: Snapshot of MIST's illustrations of Vivaldi Winter 1st Movement: Solo Violinist in high octave. MIST generates the visuals in real time in response to live music, and is best viewed in video.

present a demo of it based on a live violin performance of excerpt from Schindler's List Theme at `https://www.youtube.com/watch?v=A5plR9gs0QU`. Note how often broken hearts come up in to capture the sadness that's heard in the music and how in one of the higher octave scenes there are clouds and small umbrellas. We also provide a demo based on a recording of Vivaldi's 1st Winter Movement: `https://youtu.be/-q9VRZ8nmcg`. Note how fall leaves come up a lot in steady scenes where the octave is low and not very loud, while broken hearts come up when the scene intensifies even if it remains in a lower octave.

Colored Stars:

Tjoa (Tjoa) created a real-time spectrogram with black asterisks printing on a white background as an example to build on for further real-time music analysis[7]. We built onto it so that as music is played various colored asterisks print out depending on the audio's amplitude with a black background used to enhance the final display. We use Librosa[2] and PyAudio[8]. We only worked with a spectrogram for this version following Tjoa's work[9] so we did not need to save to a wave file to extract more features, but this limited what features we could work with to map to colors. Using ANSI text color coding, from piano to forte amplitude is divided into 4 ranges mapping to blue, magenta, yellow, and red respectively. Red represents intensity, forte, and blue represent mellowness, piano. Thus the softer the volume the warmer the colors and the less stars there are, while the louder the volume the bolder the colors and the more stars there are to represent all the layers of sound waves the system is picking up on. Figure 3 shows an image of the colored stars in response to Vivaldi's Winter 1st Movement of a segment captured of multiple instruments playing at the same time at

[7]https://musicinformationretrieval.com/ under the section "Just For Fun: Real-Time Spectrogram"

[8]https://pypi.org/project/PyAudio/

[9]https://musicinformationretrieval.com/ under the section "Just For Fun: Real-Time Spectrogram"

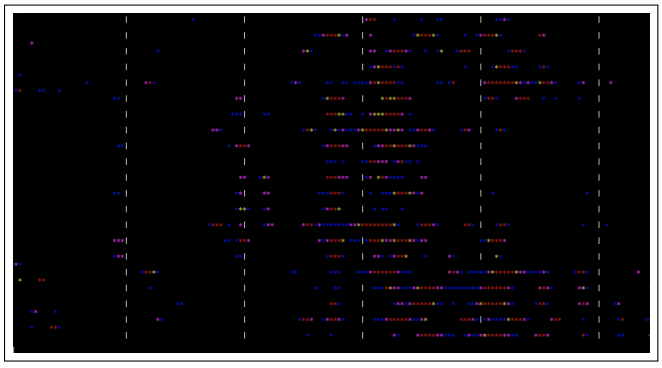

Figure 3: Snapshot of MIST's illustrations of Vivaldi Winter: Multiple Instruments.

a fast tempo[10] (Lockey 2017).

As with the Emoji version, this version of MIST also prints the sequence of symbols synchronously with the music, and is best viewed in video form. We present a demo of it based on a live violin performance of excerpt from Schindler's List Theme at https://www.youtube.com/watch?v=sFeHJ6DBXcQ. Note how the number of stars change in sync with the many changes in volume. Especially with much softer notes, the stars decrease even more such as on held-out notes where the sound fades toward the end. While, in more intense spots stars fill the screen with various colors to capture the different amplitude ranges. We also provide a demo of it based on a recording of Vivaldi's 1st Winter Movement[11] (Lockey 2017): https://youtu.be/UlANbN2IBeI. Note how in parts where multiple instruments are playing there are more starts displayed on the screen, while when the solo violinist is playing a staccato scene at 2:28 the stars are sparse and scattered.

Impact

In this section, we discuss the impact of MIST and its potential for partnering with live orchestras. We discuss how MIST can improve audience engagement and make orchestra performance accessible to wider audiences.

Strengthening Engagement

Art and emotions are closely linked (Silvia 2005). A study conducted on music-evoked emotions by examining functional magnetic resonance images proved that music evokes emotion, producing a variety of emotions depending on the music's style (Schaefer 2017). Stronger emotional connection was shown to increase attention (Tyng et al. 2017). Extracting music's attributes allows MIST to capture the music's emotions through illustrations, providing audiences audio and visual sources to connect with, which can strengthen their emotional connection.

Furthermore, interest initiates motivation which in turn increases engagement (Murayama 2018)(Tyng et al. 2017).

[10]https://www.youtube.com/watch?v=JkP7slIc9aM

[11]https://www.youtube.com/watch?v=JkP7slIc9aM

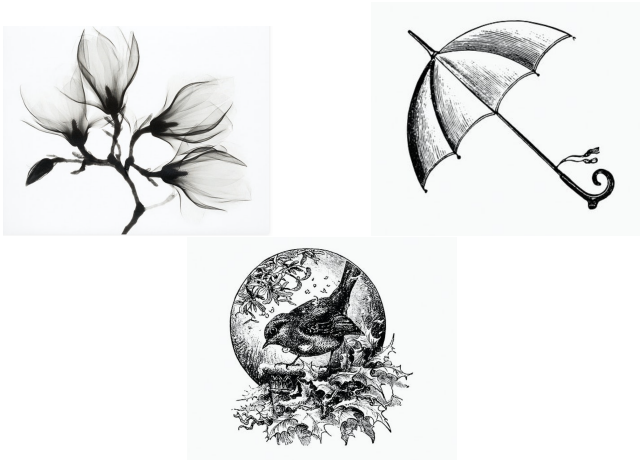

Figure 4: Images illustrating our vision for the vintage drawing style in which a mature version of MIST would be able to create scenes based on live music. Public domain images.

If the introduction of corresponding visualizations interests people, it has the potential to contribute to retaining or possibly even growing attendance at orchestra concerts. Some who may not have been interested in orchestra performances may be interested in the audio-visual experience.

There are many ways the story within a music can be visually represented. Our vision for what a mature version MIST would be able to draw are scenes illustrated using a vintage drawing style shown in Figure 4.

A New Kind Of Performance

Future versions of MIST can bring out the more complex visuals conveyed through the music. Take, for example, Vivaldi's 1st Winter Movement[12] (Lockey 2017) It opens with a lower octave evoking a feeling of suspense as it builds up the suspense reaching a higher octave that makes you feel as if you are in the midst of a storm with winds picking up. Shifts in music trigger emotion, which in turn can help you mentally visualize the music (Taruffi and Küssner 2018), (Schaefer 2017). Instead of just seeing a mental image, MIST could help showcase the story visually as a part of the performance. It has been proven seeing art has a positive impact on one's well-being. Neuroimaging studies demonstrate an immediate emotional response to artwork associated with circuitry involved in pleasure and reward (Mastandrea, Fagioli, and Biasi 2019).

Greater Access for Individuals with Hearing Disabilities

Over five percent of people, 466 million worldwide, have disabling hearing loss. (Organization 2021). By turning an orchestra's performance into illustrations that capture the music, hearing impaired individuals could understand what story an orchestra performance is sharing with its audience. A collaboration of MIST with a live performing orchestra can be a way for anyone who struggles with deafness to be able to connect with an orchestra performance. Even more so, a smaller home version of MIST could allow hearing impaired individuals to connect with any music.

[12]https://www.youtube.com/watch?v=JkP7slIc9aM

Conclusions

We offer a vision for a co-creative system that can bring to life the story conveyed through live music, proposing a new performance paradigm where live orchestras are accompanied by MIST drawing illustrations in reaction to the music. Our initial version of MIST extracts musical elements which are subsequently turned into an illustration, all in real time. Scaling the system into the full version is left for future work. Future version will consider other forms of visualization. However, due to the imaginative and divergent nature of visually interpreting music, there is a wide range of meaningful visuals for any musical piece. As such, any system would inherently represent limited visual meanings, allowing for a wide variety of approaches to this creative transformation. Furthermore, there are a multitude of approaches for extracting emotion from music (for example, in addition to the approach presented here, another method is explored in (Hevner 1936) and (Chen et al. 2008)).

In partnering MIST with live orchestras, we stand to increase audience engagement by opening two sources of information for audiences to emotionally connect with. Furthermore, we open the possibility of engaging with new audiences. This initial version considers western classical music, however, the vision presented here can be applied to a wide range of musical styles. A MIST-orchestra opens new horizons of creative engagement enabled through human-machine creative collaboration.

References

Ball, P. 2012. Iamus, classical music's computer composer, live from malaga.

Brockhoeft, T.; Petuch, J.; Bach, J.; Djerekarov, E.; Ackerman, M.; and Tyson, G. 2016. Interactive augmented reality for dance paper type: System paper.

Brown, A. R. 2018. Creative improvisation with a reflexive musical bot. *Digital Creativity* 29(1):5–18.

Chen, C.-H.; Weng, M.-F.; Jeng, S.-K.; and Chuang, Y.-Y. 2008. Emotion-based music visualization using photos. In *International Conference on Multimedia Modeling*, 358–368. Springer.

Diaz-Jerez, G. 2011. Composing with melomics: Delving into the computational world for musical inspiration. *Leonardo Music Journal Issues* 21:13–14.

Hevner, K. 1936. Experimental studies of the elements of expression in music. *American Journal of Psychology* 48:246–268.

Lockey, N. 2017. Antonio vivaldi and the sublime seasons: Sonority and texture as expressive devices in early eighteenth-century italian music. *Eighteenth Century Music* 14(2):265–283.

Mastandrea, S.; Fagioli, S.; and Biasi, V. 2019. Art and psychological well-being: Linking the brain to the aesthetic emotion. *Frontiers in Psychology* 10:739.

Murayama, K. 2018. American psychological association: The science of motivation.

Organization, W. H. 2021. Deafness and hearing loss.

Putman, A., and Keller, R. 2015. A transformational grammar framework for improvisation.

Schaefer, H.-E. 2017. Music-evoked emotions-current studies.

Silvia, P. 2005. Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology* 9:342–357.

Taruffi, L., and Küssner, M. 2018. A review of music-evoked visual mental imagery: Conceptual issues, relation to emotion, and functional outcome.

Tjoa, S. Real-time spectrogram.

Tyng, C. M.; Amin, H. U.; Saad, M. N. M.; and Malik, A. S. 2017. The influences of emotion on learning and memory. *Frontiers in Psychology* 8:1454.

Weinberg, G., and Driscoll, S. 2006. Robot-human interaction with an anthropomorphic percussionist. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, 1229–1232. New York, NY, USA: Association for Computing Machinery.

Williams, K., and Keen, D. 2008. Survey of public participation in the arts. *National Endowment for the Arts*.