INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

Vol 14, Issue 3, (2024) E-ISSN: 2222-6990

A Literature Review on Music Parameter Extraction and Visualization

[1]Deran Jin, [1,2]Mohd Zairul, [1]Sarah Abdulkareem Salih

[1]Department of Architecture, Faculty of Design and Architecture, [2]Institute for Social Science Studies, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v14-i3/21093 DOI:10.6007/IJARBSS/v14-i3/21093

Published Date: 03 March 2024

Abstract

Music visualization research is extremely complex and dynamic. Several researchers have applied various methods to persevere in the study of all aspects that make up music. The complexity of music also includes factors such as waveform, frequency, pitch, rhythm, tempo, timbre, and chords. Researchers in recent years have studied the extraction of single elements, visualization, or cross-discipline for these aspects. As far as the current research is concerned, most of the disciplines related to music visualization are focused on computers, psychology, sports science, and other related disciplines. Research on the elements of music itself has focused on music visualization, music element extraction, music association, music emotion, and the study of several important aspects of music, such as waveform, frequency, pitch, rhythm, tempo, timbre, and chord. After reviewing the research, this paper has found that with the continuous development of science and technology, music visualization has a progressive intersection with computer science, artificial intelligence, and neural networks. Thus, future research can continue to interact more with computer science.

Keywords: Music, Extraction, Visualization, Waveform, Frequency, Pitch, Rhythm, Tempo, Timbre, Chords

Introduction

With the increasing research in music visualization, music studies, visualization, element extraction, and related parameters are conducted dynamically. For music-related studies, the first is the study of association with some researchers conducting music-to-colour association studies and music-to-emotion-to-colour association studies (Zamm et al., 2013). Music-emotion correlation studies are also ongoing research in the field of music research. Many scholars have used it to study the effect of music on listeners' inter-emotional responses by studying people recording their emotional reactions after listening to music (Swaminathan & Schellenberg, 2015). Many researchers have also studied the extraction or visualization of some aspects of music through computer technology and datasets, and the study of music

also has cross-disciplinary studies with the human brain and motor nerves, among others (Roy & Dowd, 2010).

This paper reviews the methods related to music content processing in terms of element extraction and visualization. In terms of research on music element extraction, some researchers' music information retrieval algorithms have been used to extract the contours of melodies in music (Zhang, 2022). Some researchers have also used neural networks to track and extract rhythms and beats in the music and so on (Oord et al., 2018). For music visualization research, scholars have been working on developing models with neural networks to continuously improve the precision and accuracy of visualization (Miller et al., 2019; Lima et al., 2022).

Music is an extremely complex object of study in itself, and music visualization can be designed to cover all aspects of music, such as rhythm, waveform, and melody (Yu et al., 2021). The rest of the elements have been studied in the direction of how to extract them from other characteristics of the music and analyze their accuracy after extraction (Pinto et al., 2021). During the course of the review of music visualization, many studies were found to have accomplished graphical, pictorial treatments of music but failed to respond with sufficient precision to the music itself (Lima et al., 2022). Many visualization studies have visualized music from an associative perspective, and most of these visualization methods examine music as a whole (Fonteles et al., 2014). By contrast, other studies on musical elements have mostly focused on the extraction of those elements, and extraction can be considered the first step of visualization (Zhang, 2022). Therefore, this paper posits that the visualization of each element of music can still undergo further investigation on the basis of extraction. Additionally, being able to restore the characteristics of each element of music more accurately is the direction in which visualization research can continue to dig deeper. The accuracy of the methods for visualization extraction has also become more accurate with the development of the research and the aspects studied (Lima et al., 2022). With the development of the disciplines of artificial intelligence and neural networks, an increasing number of researchers are working on a more accurate representation of the visual content of music based on neural networks (Kim et al., 2019).

The challenges in music visualization are primarily reflected in the following aspects: most of the studies are focused on a specific dataset with certain limitations (Greer et al., 2019). In addition, most of the current research in music visualization happens through real-time animation, with less attention to the structural components of music (Lima et al., 2022). In terms of the current visualization research, the accuracy of the method of extracting and tracking the element can be further improved, regardless of the musical element, which is often difficult to improve in previous studies (Huang et al., 2021). As far as the current research is concerned, the visualization studies of melody and rhythm in music are well-established (Salamon et al., 2012; Reddy & Rompapas, 2021). Moreover, visualization studies for other musical elements are still more difficult (Lima et al., 2022; Reddy & Rompapas, 2021).

Music technology is constantly changing, and new important techniques are emerging to visualize music content (Miller et al., 2019). Research on music visualization helps us understand music better and more intuitively (Dalton et al., 2019). Music is also often used

as a tool for emotional expression by artists, and research on the extraction and visualization of musical elements can provide a better insight into the artist's artistic style (Coorevits et al., 2019). It also provides an easy and more accurate match for people to search for music that matches their preferences in their lives (Zhang, 2022). Music often affects people's psychological conditions and even physical motor performance; thus, visualizing or extracting music-related elements provides a better understanding of the mechanisms by which music affects people's minds and bodies (Karageorghis et al., 2018). Exploring music element extraction and visualization studies can also further inform music classification (Eghbal-Zadeh et al., 2015).

It provides a research basis and theoretical support for future researchers studying the direction of music visualization. In the study of visualization, less research is carried out on pitch, spacing, timbre, and harmony of music in the study, and these topics can exactly provide new directions for researchers. Music visualization research will provide more comprehensive theoretical support for a better understanding of music.

Materials

Music

In the field of music research, associative studies also reached a peak in 2013. Zamm et al.(2013) proposed the color-music Association (CSM) which refers to the phenomenon of perceiving color when someone hears a note or sings a song. Palmer et al.(2013) demonstrated experimental evidence for cross-modal matching between music and color mediated by emotional associations. White matter correlates of colour-music associations of synapses have also been investigated (Zamm et al., 2013). Music and emotion have been the subject of keen research, and Clarke et al (2015) have extensively studied emotional, linguistic, and social motivation from a musical perspective. A large number of academic papers and awards from different disciplines are presented, demonstrating that listening to music via headphones can profoundly change the cultural attitudes of highly demanding perceivers. Swaminathan and Schellenberg (2015) investigated the link between music and emotion, the communication and perception of emotion in music, the emotional consequences of listening to music, and the predictors of music preference. Music audio generally consists of three physical attributes: frequency, time, and amplitude (Koelsch et al., 2013).

Some researchers have demonstrated the processing of non-local dependencies in music (Koelsch et al., 2013). Li et al (2018) proposed a speech analysis dataset for facilitating musical performances and informed us how to build a complete dataset from a very small package. Melody extraction algorithms are used in computer science to extract pitch information about the main melody from music recordings (Salamon et al., 2014). Levitin et al (2018) review studies targeting the temporal and rhythmic characteristics of music that span several methodological techniques, including neurosurgery, psychophysicists, and traditional behavioural experiments. We also review studies of animal synchrony and compare the results to advances in human rhythm perception and cognition. MIR has been exploring automated music genre recognition since 2002. Our strategy based on feature sets is effective. Analyzing, evaluating, comparing, and merging acoustic and visual features produce classification accuracies equivalent to or better than existing methods (Nanni et al., 2016). Herremans et al (2018) proposed a functional classification that reveals the

interconnectedness of systems used for automatic music generation systems. Interference models, composer models, and hybrid models differ in their assumptions and network structure. Dong et al (2018) proposed three symbolic multitrack music generation models based on MuseGAN.

Some researchers reviewed the research results on content-based music information retrieval involving eight denotational-related tasks, including sound/non-sound segmentation, artist identification, style classification, dance identification, sentiment identification, instrument identification, and music fragment annotation (Murthy & Koolagudi, 2019). Flexer et al (2020) reviewed the latest breakthroughs in music structure analysis methods for audio and discussed the challenges that may arise when applying these techniques in the real world. Nieto et al (2020) explored the latest algorithms for music structure analysis of audio and their problems in real life. Calvo-Zaragoza et al (2021) examined the application of optical music recognition (OMR) to transform digital audio data into non-digital audio data. Lerch and Knees (2021) investigated the new approaches in the field of music information retrieval and audio signal processing new approaches in the field of music information retrieval and audio signal processing, mainly through machine learning solutions.
Koelsch and Jäncke(2015) proposed a new assessment method to measure heart rate changes associated with music and other factors. Some scholars (Janata et al., 2012) have experimentally explored the relationship between sound and emotion, arguing that sound can be considered a mental structure and model system. Roy and Dowd (2010) assessed the involvement of acoustic systems in terms of musical neurochemistry. McDermott et al (2016) studied the same music in different ethnic and regional cultures in terms of aesthetic responses and found that exposure to musical harmony may alter tastes, demonstrating that culture dominates aesthetic responses to music. Roy and Dowd (2010) examined how individuals and groups use music from a sociological perspective, how the collective production of music is achieved, and how music relates to broader social distinctions, particularly class, race, and gender.

Extraction
Schedl et al (2014) first introduced established methods for feature extraction and music indexing of music items from audio signals and background data sources, focusing on contemporary MIR achievements (e.g., automatic semantic tagging and user-centric retrieval and recommendation methods). Methods for estimating heterotopic/polyphonic music melodic sequences based on systematic MODGD (direct) and source-based MODGD (source) have also been investigated (Rajan et al., 2017). Chu (2022) analyzed the characteristics of digital music and extracted musical features, rhythm, tune, intensity, and timbre in MIDI format. To extract musical melodies, Bittner et al (2015) trained a discriminative binary classifier to identify melodic and non-melodic contours. It outperformed the generative model in contour classification accuracy.

Oramas et al (2016) created a music knowledge base and tested an information extraction pipeline to better interpret the music data. To demonstrate that signal acoustic features can be used to distinguish musical genres, Shin et al (2019) used a sound-encoded auditory spike code to extract acoustic features similar to the human auditory system. In the same year, MFCC was also used to extract features, and K-NN was used to classify music as pop or RnB

(rhythm and blues) (Ramadhana & Widiarthaa, 2021). In the study of cross-modal learning (e.g., audio and lyrics), Shin et al (2019) proposed a cross-modal deep associative learning architecture with a two-branch deep neural network to process audio and text (lyrics) used a multi-agent system that assigned extraction, classification and service duties, thereby enabling the automatic classification of music. On the other hand, Mo & Niu (2017) used orthogonal matching pursuit, Gabor function, and Wigner distribution function to analyze music signals—OMPGW extracts music sentiments for music applications, such as music retrieval or recommendation.

Visualization

In contrast to the music notation method (the most popular previous method for visualizing music), Smith and Williams(1997) proposed an alternative method for visualizing music using color and three-dimensional space. Cooper et al (2006) reviewed the attempts to visually represent high-level information about music content and reviewed new methods for visualizing music with MIR in the field of music visualization as the current state of the art. Nanayakkara et al (2007) quickly established real-time music visualization by using a combination of Max/MSPTM and FlashTM to propose a novel, new scheme that can be used to visualize music. Puzoń and Kosugi (2011) found that Visuals demonstrated not only obvious repetitions when reading a score or listening to a piece of music but also more subtle repetition patterns. Donnelly and Sheppard(2013) explored the use of Bayesian networks to identify the timbre of musical instruments. Bayesian networks with conditional dependencies in the time and frequency dimensions achieved 98% accuracy in the instrument classification task and 97% accuracy in the instrument family identification task.

A simplified 3D particle system and a fast translation algorithm have also been implemented to generate real-time animated particles orchestrated by classical music for music visualization (Fonteles et al., 2013). To develop a support tool for music perception and composition, Fonteles et al (2014) proposed a 3D particle system and a mapping algorithm. Lex et al (2014) used a novel visualization technique, namely, UpSet, to quantitatively analyze sets and their intersections and aggregates of intersections, to visualize music.

For children with hearing impairments, Kim et al (2015) provided a prototype of a music visualization system that records and decodes musical elements into digital data and then visualizes the information. Some scholars have analyzed the visualization elements through the ability to read the soundtrack or even simply listen to a live performance to understand the structural components of the piece (Malandrino et al., 2015). Oramas et al (2016) showed that visualization can improve the recognition of musical forms by examining the theory of isochore structure, visualization, and comments from novices and veterans. Pons & Serra (2017) used convolutional neural networks with tiny rectangular filters to classify music. A more expressive and intuitive deep learning architecture was achieved through the representational power of the first layer and the application of various filter shapes based on the musical concepts within the first layer. To help people create musical compositions quickly and efficiently, Malandrino et al (2018) built a visual tool called, Visual Harmony.

Miller et al (2019) used a fundamental concept in music theory, namely, the circle of fifths, as a model for studying visualized music. Jeong and Kim (2019) linked the DMX512 protocol through Openframeworks to create a ‘dynamic lighting for music visualization’ to present

musical features. Khulusi et al (2020) investigated and overviewed the special relationship between musicology and visualization. To reveal the semantic structure in classical orchestral works, Chan et al (2009) proposed an innovative visualization solution. Ciuha et al (2010) visualized music by interconnecting similar aspects of music and visual perception, with their research focusing on visualizing the harmonic relationship between pitch and color. To enhance the music listening experience in private spaces, Reddy and Rompapas(2021) used 'liquid hands' to bridge the distance between visualized virtual and actual concerts by utilizing alternative solutions in a virtual environment (Isaacson, n.d.). Lima et al (2022) recommended classifying ideas through input attributes, visualization quality, InfoVis technique if interactivity is allowed, and user assessment. This paper examines visual techniques developed by experts in the field of music analysis as well as some less successful approaches to music visualization Ohmi (2007) In the current work, the authors attempt to illustrate the development of music through still images of music in specific time units whilst paying attention to its structural components rather than through real-time animation.

Waveform

Inspired by sawtooth waves, Camacho and Harris (2008) developed SWIPE, an estimator for evaluating the pitch of speech and music. Klapuri (2008) determined the fundamental wavelengths of multiple sounds in polyphonic music and multichannel speech signals by studying computer models of human auditory regions. WaveNet emerged in the field of sound waveform extraction, which is a deep neural network for generating raw audio waveforms (Oord et al., 2016). Oord et al (2017) investigated probability density distillation, a novel method for training parallel feedforward networks to generate high-fidelity speech samples 20 times faster than real-time using learned WaveNet. Shen et al (2018) used a modified WaveNet model as a vocoder in combination with Tacotron 2 (a neural network structure for text-to-speech synthesis) to generate time-domain waveforms in spectrograms. Rethage et al (2018) used wavelet's end-to-end speech denoising learning method, which enabled researchers to create a model that maintains 'in-phase' signals in the waveform graph to overcome the shortcomings of amplitude spectrograms. Lluís et al (2019) developed a deep learning model based on spectrograms—DeepConvSep, which can be improved by our proposed Wavenet-based model and Wave-U-Net. Nakamura and Saruwatari (2020) proposed a governmental deep neural network based on the Wave-U-Net discrete small Wavelet Transform (DWT); DWT is used for time domain music source waveform separation. Wu et al (2021) proposed a pitch-adaptive waveform generation model called, Quasi-Periodic Wave Network (QPNet), to overcome the limited pitch controllability of fictitious wave networks (WNs) using pitch-dependent expanded convolutional neural networks (PDCNNs). In contrast to many audio synthesis efforts, where direct waveform creation models perform best, the state-of-the-art music source separation is the computational masking of the amplitude spectrum (Défossez et al., 2021). By modeling the correlation of the spectrogram along the time and frequency dimensions, Chen et al (2022) proposed a host- and network-based time-frequency attention module and multiscale attention to effectively capture the association of music signals and explore the connection between music spectrograms and music waveforms.

To enable the model to better choose whether the acoustics are in the spectral or waveform domain, Défossez et al (2021) investigated how to perform end-to-end hybrid source

separation. In recent years, song separation SVS algorithms dealing with encoder potential waveform graphs have improved in quantity and quality (Papantonakis et al., 2022).

Pitch

McLeod and Wyvill (2003) created software that can accurately display the pitch of notes being played or sung by a musician in real time. Some researchers have proposed that effective noisy speech multi-pitch tracking algorithms are essential for acoustic signal processing (Wu et al., 2003). Povey et al (2011) proposed a new speech recognition method using a Gaussian mixture model with the same number of Gaussians in all Hidden Markov Model stages, with each state having a 50-dimensional vector and a parameter to the GMM global mapping of the space.

Pitch is one of the main auditory senses and plays a decisive role in the analysis of music, speech, and auditory scenes (Oxenham, 2012). Zatorre and Baum (2012) claimed that speech and musical melodies process pitch information differently with two pitch-related processing systems, one for coarse-grained approximate analysis and one for finer-grained accurate representation, which is unique to music. For an automatic speech recognition system, Ghahremani et al (2014)proposed a method for estimating pitch and articulation probabilities. The BABEL project investigated data from multiple languages and found considerable improvements over systems without pitch features and systems that obtained pitch and POV information via SAcC or getf0. Kim et al (2018) proposed a data-driven pitch tracking algorithm, CREPE, which outperformed the test-performing PYIN algorithm technique, thus far, which is based on a deep convolutional neural network operating directly on time-domain waveforms. Huang et al (2021) proposed an RNN-based encoder-decoder framework for simulating the state cost estimation and Viterbi backtracking channels of the RAPT algorithm. Experiments on tone extraction show that the proposed tone-tracking model is better than DNN-RNN and bidirectional variants.

Hosoda et al (2021) proposed a narrowband speech pitch estimation technique using harmonic phase difference. Blok et al (2021) investigated an improved instantaneous frequency and power-based pitch estimation method, namely, IFE, to exploit the openness of pitch estimation in signal processing research. To improve the accuracy of fundamental frequency estimation, Queiroz and Coelho (2022) proposed a new method that classifies noisy speech into low- or high-frequency frames with that feature specifying the F0 frequency of the speech, classifying the frames as low or high frequency. This separation improves the F0 estimation by correcting the candidates of classical fundamental frequency detection methods.

The proposed technique outperforms existing solutions in terms of low/high-frequency separation accuracy. Bittner et al (2022) introduced a lightweight neural network for instrument transcription that supports multiple vocal outputs, which can be extended to many instruments (including the human voice) in this study. The multi-output structure of our model increases frame-level note accuracy by simultaneously predicting frame-level onsets, multiple pitches, and note activations. To control the extraction of attention-fused vocal melodies, Yu et al (2023) developed a neural harmonic perception network. Papantonakis et al (2022)investigated the effect of visual feedback on the ability to recognize and consolidate pitch information.

Rhythm

Beginning with psychophysical studies of temporal rhythm and pitch perception, Krumhansl (2000) summarised psychological research on how this aspect is seen and recalled. Patterns, beats, and rhythms are the temporal components of rhythm. Music rhythmically activates the somatic and premotor systems (Thaut et al., 2014). By studying percussion, Repp (2005) showed that sensorimotor synchronization (SMS), the rhythmic coordination of perception and action, is most evident in music and dance. For studies of rhythmically responsive motor areas, Grahn and Brett (2007) suggested that basal ganglia and SMAs may mediate rhythmic perception outside of motor creation. Some scientists have also used rhythmic features to create a Thayers-based model of emotion to investigate the association between emotion and rhythm (Cu et al., 2012). Böck et al (2016) provided a then state-of-the-art method for extracting combined beats and low-tempo rhythms from audio sources. To provide music mixing with rhythmic synchronization, extraction of rhythmic patterns, and rhythm-based music retrieval, Lin et al (2010) developed methods that automatically select similar songs by a seed song and user-defined rhythmic parameters. Quinton (2017) evaluated the reliability of rhythmic feature extraction to improve the confidence of automatic beat structure analysis and MIR systems.

The study provides two methods to automatically quantify metric modulation in audio recordings. Automatically ‘capturing rhythms’ and annotating musical beats to correct them have been a topic faced by scientists(Driedger et al., 2019). Driedger et al (2019) provided a novel dataset displaying beats and mathematically describing the automatic correction method, demonstrating its effectiveness. Dalton et al (2019) explored how rhythm analysis enables DAW rhythms to synchronize with source recordings. Research by Percival and Tzanetakis inspired Renoise's basic beat extraction technique. Böck and Davies (2020) evaluated cutting-edge deep neural network techniques for computational rhythm analysis, which improved the performance of the system by 6% by disassembling, examining, and reassembling such techniques.

Chords

The EEG of music cognition has rarely been studied (ERPs); Koelsch et al (2000) determined that musical context, task relevance of accidental chords, degree of violation, and probability influenced music processing. Pauwels & Peeters (2013) provided a new approach to music structure segmentation based on an integrated estimate of structural segments, keys, and chords in a probabilistic framework. A priori probabilities of key changes and chord transitions define the boundaries of the structural segments. To investigate neonatal responses to Western music, Virtala et al (2013) tested change-related mismatch responses (MMR) by encoding Western music chords in the neonatal brain using ERPs.

Virtala et al (2014) study determined that musicology improves brain and behavioral recognition of Western music chords. Cambouropoulos et al (2014) study found that an idiom-independent chord type representation captured tonal simultaneity in every harmonic context, leading them to focus on harmonic representation and computational analysis (e.g., modal, modal, jazz, octave, and atonal). To explore chords with various affective properties, the work of Lahdelma and Eerola (2016) examined the affective nature of vertical harmonies. To anticipate the feelings of listeners of musical parts, Greer et al (2019) investigated a corpus of chords and lyrics matched to musical phrases, which were used to represent lyrics and

chords in a shared vector space. To enable users to turn images into short chord-spin combinations, Polo and Sevillano (2019) developed Musical Vision, an emerging tool that interacts to construct fully variable mappings between color space and MIDI instrument and audio pitch space.

Melody

Polansky & Bassein(1992) used contour theory to assess pitch means in large-scale segmentation of waveforms, melodies, musical pieces, or other measurable features. Halpern and Zatorre (1999) used PET to examine brain activity associated with known melodies. Margulis (2005) provided an empirical technique for analyzing melodic anticipation and a model for rating the anticipatory nature of the melodic occurrence. Salamon and Gómez (2012) developed a unique method for extracting major melodies from polyphonic music recordings. In the same year, Salamon et al (2012) provided a unique method for genre identification by directly exploiting the high-level melodic qualities in the audio signal of polyphonic music. Later, Salamon et al(2014) summarised the difficulties in the design, evaluation, and application of melody extraction methods and proposed that melody extraction research faces problems in algorithm performance, development, and evaluation. Zhang (2022) proposed a LAM algorithm based on music melody contour feature extraction and oriented to music information retrieval.

Beats

Ariza & Cuthbert (2010) applied the beat module of the TimeSignature-music21 Python toolbox to read Humdrum and MusicXML and output Lilypond and MusicXML. Salamon et al (2012) found that beat perception is one of the auditory input recognition regular pulses that may contribute to the creation of music. The findings clearly support intrinsic beat perception. Degara et al (2011) used a novel probabilistic approach to calculate the duration between musical beat occurrences by explicitly modeling non-beat states. Holzapfel et al (2012) proposed a beat technique for tracking difficult musical samples without ground truth. Böck et al (2019) proposed a multi-task learning method for musical rhythm estimation and beat tracking trained entirely using rhythmic annotations. Böck et al (2019) used a new method of temporal convolutional networks to monitor audio in beats. In the same year, the researchers also designed a causal technique for determining the location of beats from an audio source (Richter, 2019).

The data-driven automatic drumming transcription (ADT) model of Wang et al (2020) was unable to discriminate beats outside of a specified, small range of percussion-like vocabularies. The ADT problem for open vocabularies was overcome by adding a bit of learning. To eliminate the designed spectral features, Steinmetz and Reiss (2021) developed WaveBeat, a waveform-based end-to-end joint beat and downbeat tracking method. Pinto et al (2021) proposed a real-world beat-tracking strategy based on a relatively small temporal region of annotated beat positions and focused fine-tuning of the most advanced deep neural network to extract beats from music audio signals. To improve the accuracy and relevance of beat matching, Zhu (2022) developed a data mining-based error recognition system for dance movement and music beat matching.

Tempo

To measure the effect of music on the assessment of happiness and depression, Bella et al (2001) concluded that tempo mastery precedes modality when interpreting the emotional tone conveyed by music. Karageorghis et al (2008)examined how music rhythm influences motor flow, intrinsic motivation, and music choice. To further understand the connection between music and emotion, Van Der Zwaag et al (2011) investigated how rhythm, pattern, and tempo influence mood. To test linear and nonlinear models for predicting musical tension, Farbood (2012) examined several musical factors: harmony, pitch, melodic anticipation, dynamics, onset frequency, tempo, beat, rhythmic regularity, and syncopation.

Getz et al (2014) again studied to assess how stress, optimism, and musical training affect a person's desire to listen to music (for emotional control and/or cognitive stimulation) and the tempos they prefer. As the research on music and emotion deepens, some researchers believe in the importance of using music as a pre-game technique in sports by adjusting the volume and tempo of music whilst monitoring brain activity (Bishop et al., 2014). Another approach offered by Juslin et al (2014) attempted to explain musical emotions in terms of a set of target mechanisms triggered by various information in musical events (e.g., tempo). Percival & Tzanetakis (2014) proposed a reduced tempo estimation method for music with constant or near-constant tempo to retain tempo accuracy whilst reducing steps, parameters, and modeling assumptions. Building on previous research on tempo-emotion associations, Dobrota & Reić Ercegovac (2015) investigated topics aimed at understanding whether a correlation exists between listeners' preferred patterns and tempos and their individual personality attributes.

Rosemann et al (2016)studied the effects of eye-hand coordination, performance tempo, complexity, and cognitive abilities of pianists. Neuhoff et al (2017) explored the challenges in tempo playing methods, variations in fine-tuning and expressiveness, temporal effects, and the implications of these results for music theory. Karageorghis et al (2018) assessed the interactive effects of musical tempo and intensity (volume) on the execution and subjective emotion of a basic motor skill by assessing whether this study further extends previous research. Coorevits et al (2019) again returned to the musical performance itself, examining the many effects that changes in musical tempo have on the ‘performance state’ or the articulation of the performer's movements. Bittner et al (2022) offered a way to understand musical tempo beyond listening to music by developing a Visual project and found that the number of notes per time unit and tempo also mattered.

Timbre

Pressnitzer et al (2000) concluded from a psychoacoustic perspective that psychoacoustic roughness increases when non-tonal orchestral timbres reduce musical tension. Patil et al (2012) studied musical timbres using a neurocomputational framework with a nonlinear classifier and 1,000 mammalian primary auditory cortical neurons and spectral-temporal receptive areas of simulated cortical neurons.

Burger et al (2013) simultaneously investigated the relationship between rhythm, timbre, and motion, concluding that body motion reflects, reproduces, and predicts musical quality. Town and Bizley (2013) outlined human timbre perception and the spectral and temporal acoustic features that shape timbre in speech, music, and environmental sounds, suggesting some

worthwhile directions for research. Lui (2013) developed a technique for teaching musical timbre on mobile devices, and the model was self-trained by volume-tuning streamlined spectral data. To confirm that the pitch and timbre analysis process of music has unexpected similarities, Cousineau et al (2014) explored music with a sequence discrimination task. Rocha et al (2013) further investigated musical similarity, focusing on electronic dance music (EDM) using timbre similarity as a sub-similarity. To overcome the drawback that both operations in modulation analysis may erase useful modulation information, Ren et al (2015) proposed a two-dimensional representation of acoustic and modulation frequencies to extract joint features.

Still, from the perspective of music retrieval classification, Eghbal-Zadeh et al (2015) used music timbre similarity and music i-vectors to derive song-level descriptors from frame-level temporal information for artist classification. Lu et al (2019) developed MUNIT to enable music classification, developed MUNIT for multimodal music style transformation and timbre enhancement using unsupervised, non-parallel data describing the multimodal scattering of musical situations. Kim et al (2019) created a neural music synthesis model with configurable timbres using sheet music and instrument data, using conditions for learning instrument embedding and WaveNet vocoder for the recurrent neural network.

Hypothetical Future and Recommendation

A review of research on music visualization found that for the aspect of music waveform research, a steady stream of researchers has emerged in recent years to study the connection between music graphs and music waveforms based on WaveNet. Pitch is also an integral part of music. In recent years, researchers have been using neural networks to model and study pitch in music. Rhythm is often associated with movement, and some scholars wish to study the effect of rhythm on motor performance, whilst others have classified music by extracting rhythmic features. Harmonic spins, on the other hand, are more linked to brain science. Research on the visualization of harmonic spins has focused more on the association with color. My interest in melody has been in melody extraction and prediction of melodic occurrence. The latest research is on the extraction of musical contour features by melody for the classification of music. The study of beats has been focused on the perception, tracking, and extraction of beats. The study of rhythm has focused more on the relationship between rhythm and motor performance. The study of tone has focused more on the area related to music classification.

Through the above studies, we found that the waveforms, harmonies, and melodies are more closely related to visualization studies of music. Visualizing waveforms and melodic colors is easier. The main limitation is that the visualization of rhythm in music is more difficult, and most of the studies have focused on the effect of rhythm or beat on motion performance. The visualization of pitch and timbre has also received less attention, and the studies are more closely related to music feature retrieval and classification.

Novelty

In the future, with the continuous development of artificial intelligence in computer science, other researchers will use new models and new algorithms to extract the characteristics of the parameters of music and further research on how to use these new technologies in improving the study of music visualization.

Reference

Ariza, C., & Cuthbert, M. S. (2010). Modeling Beats, Accents, Beams, And Time Signatures Hierarchically With Music21 Meter Objects. *ICMC*. https://www.academia.edu/download/6799610/meterobjects.pdf

Bishop, D. T., Wright, M. J., & Karageorghis, C. I. (2014). The tempo and intensity of pre-task music modulate neural activity during reactive task performance. *Psychology of Music*, *42*(5), 714–727. https://doi.org/10.1177/0305735613490595

Bittner, R. M., Bosch, J. J., Rubinstein, D., Meseguer-Brocal, G., & Ewert, S. (2022). A Lightweight Instrument-Agnostic Model for Polyphonic Note Transcription and Multipitch Estimation. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 781–785. https://doi.org/10.1109/ICASSP43922.2022.9746549

Bittner, R. M., Salamon, J., Essid, S., & Bello, J. P. (2015). Melody extraction by contour classification. *International Conference on Music Information Retrieval (ISMIR)*. https://hal.science/hal-02943532/

Blok, M., Banaś, J., & Pietrolaj, M. (2021). IFE: NN-aided Instantaneous Pitch Estimation. *2021 14th International Conference on Human System Interaction (HSI)*, 1–7. https://doi.org/10.1109/HSI52170.2021.9538713

Böck, S., & Davies, M. E. (2020). Deconstruct, Analyse, Reconstruct: How to improve Tempo, Beat, and Downbeat Estimation. *ISMIR*, 574–582. https://program.ismir2020.net/static/final_papers/223.pdf

Böck, S., Davies, M. E., & Knees, P. (2019). Multi-Task Learning of Tempo and Beat: Learning One to Improve the Other. *ISMIR*, 486–493. https://archives.ismir.net/ismir2019/paper/000058.pdf

Böck, S., Krebs, F., & Widmer, G. (2016). Joint Beat and Downbeat Tracking with Recurrent Neural Networks. *ISMIR*, 255–261. https://archives.ismir.net/ismir2016/paper/000186.pdf

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of Rhythm- and Timbre-Related Musical Features on Characteristics of Music-Induced Movement. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00183

Calvo-Zaragoza, J., Jr., J. H., & Pacha, A. (2021). Understanding Optical Music Recognition. *ACM Computing Surveys*, *53*(4), 1–35. https://doi.org/10.1145/3397499

Camacho, A., & Harris, J. G. (2008). A sawtooth waveform-inspired pitch estimator for speech and music. *The Journal of the Acoustical Society of America*, *124*(3), 1638–1652.

Cambouropoulos, E., Kaliakatsos-Papakostas, M. A., & Tsougras, C. (2014). An idiom-independent representation of chords for computational music analysis and generation. *ICMC*. http://users.auth.gr/~emilios/papers/icmc-smc2014-GCT.pdf

Chan, W.-Y., Qu, H., & Mak, W.-H. (2009). Visualizing the semantic structure in classical music works. *IEEE Transactions on Visualization and Computer Graphics*, *16*(1), 161–173.

Chen, L., Zheng, X., Zhang, C., Guo, L., & Yu, B. (2022). Multi-scale temporal-frequency attention for music source separation. *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. https://ieeexplore.ieee.org/abstract/document/9859957/?casa_token=GlxmE9jMDAsAAAAA:ALnSmw-XJdKNYwxmTb10Y12XcYrEwKTstEjNrq6l4TQr6LYboINPWQwom46zxrl643ALXCXNRWE

Chu, X. (2022). Feature Extraction and Intelligent Text Generation of Digital Music. *Computational Intelligence and Neuroscience*, *2022*.

https://www.hindawi.com/journals/cin/2022/7952259/

Ciuha, P., Klemenc, B., & Solina, F. (2010). Visualization of concurrent tones in music with colors. *Proceedings of the 18th ACM International Conference on Multimedia*, 1677–1680. https://doi.org/10.1145/1873951.1874320

Clarke, E., DeNora, T., & Vuoskoski, J. (2015). Music, empathy, and cultural understanding. *Physics of Life Reviews*, *15*, 61–88.

Cooper, M., Foote, J., Pampalk, E., & Tzanetakis, G. (2006). Visualization in audio-based music information retrieval. *Computer Music Journal*, *30*(2), 42–62.

Coorevits, E., Moelants, D., Maes, P.-J., & Leman, M. (2019). Exploring the effect of tempo changes on violinists' body movements. *Musicae Scientiae*, *23*(1), 87–110. https://doi.org/10.1177/1029864917714609

Cousineau, M., Carcagno, S., Demany, L., & Pressnitzer, D. (2014). What is a melody? On the relationship between pitch and brightness of timbre. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00127

Cu, J., Cabredo, R., Legaspi, R., & Suarez, M. T. (2012). On Modelling Emotional Responses to Rhythm Features. In P. Anthony, M. Ishizuka, & D. Lukose (Eds.), *PRICAI 2012: Trends in Artificial Intelligence* (Vol. 7458, pp. 857–860). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-32695-0_85

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Dalton, B., Johnson, D., & Tzanetakis, G. (2019).Daw integratedd beat tracking for music production. *Proc. Sound Music Comput. Conf*, 7–11. https://smc2019.uma.es/articles/P1/P1_01_SMC2019_paper.pdf

Défossez, A., Usunier, N., Bottou, L., & Bach, F. (2021). *Music Source Separation in the Waveform Domain* (arXiv:1911.13254). arXiv. http://arxiv.org/abs/1911.13254

Degara, N., Rúa, E. A., Pena, A., Torres-Guijarro, S., Davies, M. E., & Plumbley, M. D. (2011). Reliability-informed beat tracking of musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(1), 290–301.

Dobrota, S., & Reić Ercegovac, I. (2015). The relationship between music preferences of different modes and tempo and personality traits – implications for music pedagogy. *Music Education Research*, *17*(2), 234–247. https://doi.org/10.1080/14613808.2014.933790

Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., & Yang, Y.-H. (2018). Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://ojs.aaai.org/index.php/AAAI/article/view/11312

Donnelly, P. J., & Sheppard, J. W. (2013). Classification of musical timbre using Bayesian networks. *Computer Music Journal*, *37*(4), 70–86.

Driedger, J., Schreiber, H., de Haas, W. B., & Müller, M. (2019). Towards Automatically Correcting Tapped Beat Annotations for Music Recordings. *ISMIR*, 200–207. https://www.academia.edu/download/79136113/000022.pdf

Eghbal-Zadeh, H., Lehner, B., Schedl, M., & Widmer, G. (2015). I-Vectors for Timbre-Based Music Similarity and Music Artist Classification. *ISMIR*, 554–560. https://archives.ismir.net/ismir2015/paper/000128.pdf

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, *29*(4), 387–428.

Flexer, A., Levé, F., Peeters, G., & Urbano, J. (2020). Introduction to the Special Collection" 20th Anniversary of ISMIR". *Trans. Int. Soc. Music. Inf. Retr.*, *3*(1), 218–220.

Fonteles, J. H., Rodrigues, M. A. F., & Basso, V. E. (2014). Real-time animations of virtual fountains based on a particle system for visualizing the musical structure. *2014 XVI Symposium on Virtual and Augmented Reality*, 171–180. https://ieeexplore.ieee.org/abstract/document/6913091/

Fonteles, J. H., Rodrigues, M. A. F., & Basso, V. E. D. (2013). Creating and evaluating a particle system for music visualization. *Journal of Visual Languages & Computing*, *24*(6), 472–482.

Getz, L. M., Marks, S., & Roy, M. (2014). The influence of stress, optimism, and music training on music uses and preferences. *Psychology of Music*, *42*(1), 71–85. https://doi.org/10.1177/0305735612456727

Ghahremani, P., BabaAli, B., Povey, D., Riedhammer, K., Trmal, J., & Khudanpur, S. (2014). A pitch extraction algorithm tuned for automatic speech recognition. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2494–2498. https://doi.org/10.1109/ICASSP.2014.6854049

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience*, *19*(5), 893–906.

Greer, T., Singla, K., Ma, B., & Narayanan, S. (2019). Learning shared vector representations of lyrics and chords in music. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3951–3955. https://ieeexplore.ieee.org/abstract/document/8683735/?casa_token=YEdIbA5_CNAAAAAA:ByqvQ2TnFiBvtSZloSkf5T-XkUclouGmO1nJb4BjVHnanvIgxYKJvptG-XufCTNHVRYSPfKNRts

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*(7), 697–704.

Herremans, D., Chuan, C.-H., & Chew, E. (2018). A Functional Taxonomy of Music Generation Systems. *ACM Computing Surveys*, *50*(5), 1–30. https://doi.org/10.1145/3108242

Holzapfel, A., Davies, M. E., Zapata, J. R., Oliveira, J. L., & Gouyon, F. (2012). Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(9), 2539–2548.

Hosoda, Y., Kawamura, A., & Iiguni, Y. (2021). Pitch Estimation Algorithm for Narrowband Speech Signal using Phase Differences between Harmonics. *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 920–925.

Huang, H., Wang, K., Hu, Y., & Li, S. (2021). Encoder-decoder-based pitch tracking and joint model training for Mandarin tone classification. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6943–6947. https://ieeexplore.ieee.org/abstract/document/9413888/?casa_token=TK1jwxpmfosAAAAA:2gJRIMhWOe4EASZ8W30G3TnKTBK0vu_fbvFEbjpHlZAQGVAeXMxFZYxsqW_FHm6KP6dOx9yH2AY

Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, *141*(1), 54.

Jeong, W.-U., & Kim, S.-H. (2019). Synesthesia Visualization of Music Waveform:'Kinetic Lighting for Music Visualization. *International Journal of Asia Digital Art and Design Association*, *23*(2), 22–27.

Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, *42*(4), 599–623. https://doi.org/10.1177/0305735613484548

Karageorghis, C. I., Cheek, P., Simpson, S. D., & Bigliassi, M. (2018). Interactive effects of music tempi and intensities on grip strength and subjective effect. *Scandinavian Journal of Medicine & Science in Sports*, *28*(3), 1166–1175. https://doi.org/10.1111/sms.12979

Karageorghis, C., Jones, L., & Stuart, D. (2008). Psychological Effects of Music Tempi during Exercise. *International Journal of Sports Medicine*, *29*(7), 613–619. https://doi.org/10.1055/s-2007-989266

Khulusi, R., Kusnick, J., Meinecke, C., Gillmann, C., Focht, J., & Jänicke, S. (2020). A Survey on Visualizations for Musical Data. *Computer Graphics Forum*, *39*(6), 82–110. https://doi.org/10.1111/cgf.13905

Kim, J., Ananthanarayan, S., & Yeh, T. (2015). Seen music: Ambient music data visualization for children with hearing impairments. *Proceedings of the 14th International Conference on Interaction Design and Children*, 426–429. https://doi.org/10.1145/2771839.2771870

Kim, J. W., Bittner, R., Kumar, A., & Bello, J. P. (2019). Neural music synthesis for flexible timbre control. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 176–180. https://ieeexplore.ieee.org/abstract/document/8683596/?casa_token=SysTsnxUyiwAAAAA:nenIkKjNeykOixUX3v5KHNm5JShJ2FqznqpKJ0YPXTLRWyc8zjK-AKwFMnd2_p2gHE_ZbPcqJVg

Kim, J. W., Salamon, J., Li, P., & Bello, J. P. (2018). Crepe: A Convolutional Representation for Pitch Estimation. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 161–165. https://doi.org/10.1109/ICASSP.2018.8461329

Klapuri, A. (2008). Multipitch analysis of polyphonic music and speech signals using an auditory model. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(2), 255–266.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541.

Koelsch, S., & Jäncke, L. (2015). Music and the heart. *European Heart Journal*, *36*(44), 3043–3049.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences*, *110*(38), 15443–15448. https://doi.org/10.1073/pnas.1300272110

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*(1), 159.

Lahdelma, I., & Eerola, T. (2016). Single chords convey distinct emotional qualities to both naïve and expert listeners. *Psychology of Music*, *44*(1), 37–54. https://doi.org/10.1177/0305735614552006

Lerch, A., & Knees, P. (2021). Machine learning applied to music/audio signal processing. In *Electronics* (Vol. 10, Issue 24, p. 3077). MDPI. https://www.mdpi.com/2079-9292/10/24/3077

Levitin, D. J., Grahn, J. A., & London, J. (2018). The Psychology of Music: Rhythm and Movement. *Annual Review of Psychology*, *69*(1), 51–75. https://doi.org/10.1146/annurev-psych-122216-011740

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992.

Li, B., Liu, X., Dinesh, K., Duan, Z., & Sharma, G. (2018). Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. *IEEE Transactions on Multimedia*, *21*(2), 522–535.

Lima, H. B., Santos, C. G. R. D., & Meiguins, B. S. (2022). A Survey of Music Visualization Techniques. *ACM Computing Surveys*, *54*(7), 1–29. https://doi.org/10.1145/3461835

Lin, Q., Lu, L., Weare, C., & Seide, F. (2010). Music rhythm characterization with application to workout-mix generation. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 69–72. https://ieeexplore.ieee.org/abstract/document/5496203/?casa_token=1cJI1537zmEAAAAA:jKnVsFY_UMKHG3helg4OXLm7TALSOXonut0eCQAYzqbR_xZSDIJdQpmGrsCWRlTTKI6Ebe4L8hs

Lluís, F., Pons, J., & Serra, X. (2019). *End-to-end music source separation: Is it possible in the waveform domain?* (arXiv:1810.12187). arXiv. http://arxiv.org/abs/1810.12187

Lu, C.-Y., Xue, M.-X., Chang, C.-C., Lee, C.-R., & Su, L. (2019). Play as you like: Timbre-enhanced multi-modal music style transfer. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 1061–1068. https://aaai.org/ojs/index.php/AAAI/article/view/3897

Lui, S. (2013). A MUSIC TIMBRE SELF-TRAINING TOOL ON MOBILE DEVICE USING VOLUME NORMALIZED SIMPLIFIED SPECTRAL INFORMATION. *ICMC*. https://www.researchgate.net/profile/Simon-Lui-2/publication/288201909_A_music_timbre_self-training_tool_on_mobile_device_using_volume_normalized_simplified_spectral_information/links/580097d908aec5444b724df8/A-music-timbre-self-training-tool-on-mobile-device-using-volume-normalized-simplified-spectral-information.pdf

Malandrino, D., Pirozzi, D., Zaccagnino, G., & Zaccagnino, R. (2015). A color-based visualization approach to understand harmonic structures of musical compositions. *2015 19th International Conference on Information Visualisation*, 56–61. https://ieeexplore.ieee.org/abstract/document/7272579/

Malandrino, D., Pirozzi, D., & Zaccagnino, R. (2018). Visualization and music harmony: Design, implementation, and evaluation. *2018 22nd International Conference Information Visualisation (IV)*, 498–503. https://ieeexplore.ieee.org/abstract/document/8564210/

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception*, *22*(4), 663–714.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550.

McLeod, P., & Wyvill, G. (2003). Visualization of musical pitch. *Proceedings Computer Graphics International 2003*, 300–303. https://ieeexplore.ieee.org/abstract/document/1214486/

Miller, M., Bonnici, A., & El-Assady, M. (2019). Augmenting Music Sheets with Harmonic Fingerprints. *Proceedings of the ACM Symposium on Document Engineering 2019*, 1–10. https://doi.org/10.1145/3342558.3345395

Mingyang Wu, DeLiang Wang, & Brown, G. J. (2003). A multipitch tracking algorithm for noisy speech. *IEEE Transactions on Speech and Audio Processing*, *11*(3), 229–241. https://doi.org/10.1109/TSA.2003.811539

Mo, S., & Niu, J. (2017). A novel method based on OMPGW method for feature extraction in automatic music mood classification. *IEEE Transactions on Affective Computing*, *10*(3), 313–324.

Murthy, Y. V. S., & Koolagudi, S. G. (2019). Content-Based Music Information Retrieval (CB-MIR) and Its Applications toward the Music Industry: A Review. *ACM Computing Surveys*, *51*(3), 1–46. https://doi.org/10.1145/3177849

Nakamura, T., & Saruwatari, H. (2020). Time-Domain Audio Source Separation Based on Wave-U-Net Combined with Discrete Wavelet Transform. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 386–390. https://doi.org/10.1109/ICASSP40776.2020.9053934

Nanayakkara, S. C., Taylor, E., Wyse, L., & Ong, S. H. (2007). Towards building an experiential music visualizer. *2007 6th International Conference on Information, Communications & Signal Processing*, 1–5. https://ieeexplore.ieee.org/abstract/document/4449609/

Nanni, L., Costa, Y. M., Lumini, A., Kim, M. Y., & Baek, S. R. (2016). Combining visual and acoustic features for music genre classification. *Expert Systems with Applications*, *45*, 108–117.

Neuhoff, H., Polak, R., & Fischinger, T. (2017). Perception and evaluation of timing patterns in drum ensemble music from Mali. *Music Perception: An Interdisciplinary Journal*, *34*(4), 438–451.

Nieto, O., Mysore, G. J., Wang, C., Smith, J. B., Schlüter, J., Grill, T., & McFee, B. (2020). Audio-Based Music Structure Analysis: Current Trends, Open Challenges, and Applications. *Trans. Int. Soc. Music. Inf. Retr.*, *3*(1), 246–263.

Ohmi, K. (2007). Music Visualization in Style and Structure. *Journal of Visualization*, *10*(3), 257–258. https://doi.org/10.1007/BF03181691

Oord, A., Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., Driessche, G., Lockhart, E., Cobo, L., & Stimberg, F. (2018). Parallel wavenet: Fast high-fidelity speech synthesis. *International Conference on Machine Learning*, 3918–3926. https://proceedings.mlr.press/v80/oord18a.html

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio* (arXiv:1609.03499). arXiv. https://doi.org/10.48550/arXiv.1609.03499

Oord, A. van den, Li, Y., Babuschkin, I., Simonyan, K., Vinyals, O., Kavukcuoglu, K., Driessche, G. van den, Lockhart, E., Cobo, L. C., Stimberg, F., Casagrande, N., Grewe, D., Noury, S., Dieleman, S., Elsen, E., Kalchbrenner, N., Zen, H., Graves, A., King, H., … Hassabis, D. (2017). *Parallel WaveNet: Fast High-Fidelity Speech Synthesis* (arXiv:1711.10433). arXiv. https://doi.org/10.48550/arXiv.1711.10433

Oramas, S., Espinosa-Anke, L., Sordo, M., Saggion, H., & Serra, X. (2016). Information extraction for knowledge base construction in the music domain. *Data & Knowledge Engineering*, *106*, 70–83.

Oxenham, A. J. (2012). Pitch Perception. *Journal of Neuroscience*, *32*(39), 13335–13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, *110*(22), 8836–8841. https://doi.org/10.1073/pnas.1212562110

Papantonakis, P., Garoufis, C., & Maragos, P. (2022). Multi-band Masking for Waveform-based Singing Voice Separation. *2022 30th European Signal Processing Conference (EUSIPCO)*, 249–253. https://ieeexplore.ieee.org/abstract/document/9909713/

Patil, K., Pressnitzer, D., Shamma, S., & Elhilali, M. (2012). Music in our ears: The biological bases of musical timbre perception. *PLoS Computational Biology*, *8*(11), e1002759.

Pauwels, J., & Peeters, G. (2013). Segmenting music through the joint estimation of keys, chords and structural boundaries. *Proceedings of the 21st ACM International Conference on Multimedia*, 741–744. https://doi.org/10.1145/2502081.2502193

Percival, G., & Tzanetakis, G. (2014). Streamlined tempo estimation based on autocorrelation and cross-correlation with pulses. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(12), 1765–1776.

Pérez-Marcos, J., Jiménez-Bravo, D. M., De Paz, J. F., Villarrubia González, G., López, V. F., & Gil, A. B. (2020). Multi-agent system application for music features extraction, meta-classification and context analysis. *Knowledge and Information Systems*, *62*(1), 401–422. https://doi.org/10.1007/s10115-018-1319-2

Pinto, A. S., Böck, S., Cardoso, J. S., & Davies, M. E. (2021). User-driven fine-tuning for beat tracking. *Electronics*, *10*(13), 1518.

Polansky, L., & Bassein, R. (1992). Possible and impossible melody: Some formal aspects of contour. *Journal of Music Theory*, *36*(2), 259–284.

Polo, A., & Sevillano, X. (2019). Musical Vision: An interactive bio-inspired sonification tool to convert images into music. *Journal on Multimodal User Interfaces*, *13*(3), 231–243. https://doi.org/10.1007/s12193-018-0280-4

Pons, J., & Serra, X. (2017). Designing efficient architectures for modeling temporal features with convolutional neural networks. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2472–2476. https://doi.org/10.1109/ICASSP.2017.7952601

Povey, D., Burget, L., Agarwal, M., Akyazi, P., Kai, F., Ghoshal, A., Glembek, O., Goel, N., Karafiát, M., Rastrow, A., Rose, R. C., Schwarz, P., & Thomas, S. (2011). The subspace Gaussian mixture model—A structured model for speech recognition. *Computer Speech & Language*, *25*(2), 404–439. https://doi.org/10.1016/j.csl.2010.06.003

Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics*, *62*(1), 66–80. https://doi.org/10.3758/BF03212061

Puzoń, B., & Kosugi, N. (2011). Extraction and visualization of the repetitive structure of music in acoustic data: Misual project. *Proceedings of the 13th International Conference on Information Integration and Web-Based Applications and Services*, 152–159. https://doi.org/10.1145/2095536.2095563

Queiroz, A., & Coelho, R. (2022). Noisy Speech Based Temporal Decomposition to Improve Fundamental Frequency Estimation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 2504–2513. https://doi.org/10.1109/TASLP.2022.3190670

Quinton, E. (2017). *Towards the Automatic Analysis of Metric Modulations* [PhD Thesis, Queen Mary University of London]. https://qmro.qmul.ac.uk/xmlui/handle/123456789/25936

Rajan, R., Misra, M., & Murthy, H. A. (2017). Melody extraction from music using modified group delay functions. *International Journal of Speech Technology*, *20*(1), 185–204. https://doi.org/10.1007/s10772-017-9397-1

Ramadhana, Z. H. G., & Widiarthaa, I. M. (n.d.). Classification of Pop And RnB (Rhythm And Blues) Songs With MFCC Feature Extraction And K-NN Classifier. *Jurnal Elektronik Ilmu Komputer Udayana P-ISSN*, *2301*, 5373.

Reddy, G. S. R., & Rompapas, D. (2021). Liquid Hands: Evoking Emotional States via Augmented Reality Music Visualizations. *ACM International Conference on Interactive Media Experiences*, 305–310. https://doi.org/10.1145/3452918.3465496

Ren, J.-M., Wu, M.-J., & Jang, J.-S. R. (2015). Automatic music mood classification based on timbre and modulation features. *IEEE Transactions on Affective Computing*, *6*(3), 236–246.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. https://doi.org/10.3758/BF03206433

Richter, J. (2019). *Style-Specific Beat Tracking with Deep Neural Networks*. https://www.static.tu.berlin/fileadmin/www/10002020/Dokumente/Abschlussarbeiten/Richter_MasA.pdf

Rocha, B., Bogaards, N., & Honingh, A. (2013). *Segmentation and timbre-and rhythm-similarity in Electronic Dance Music*. https://eprints.illc.uva.nl/482/

Rosemann, S., Altenmüller, E., & Fahle, M. (2016). The art of sight-reading: Influence of practice, playing tempo, complexity and cognitive skills on the eye–hand span in pianists. *Psychology of Music*, *44*(4), 658–673. https://doi.org/10.1177/0305735615585398

Roy, W. G., & Dowd, T. J. (2010). What Is Sociological about Music? *Annual Review of Sociology*, *36*(1), 183–203. https://doi.org/10.1146/annurev.soc.012809.102618

Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(6), 1759–1770.

Salamon, J., Gómez, E., Ellis, D. P., & Richard, G. (2014). Melody extraction from polyphonic music signals: Approaches, applications, and challenges. *IEEE Signal Processing Magazine*, *31*(2), 118–134.

Salamon, J., Rocha, B., & Gómez, E. (2012). Musical genre classification using melody features extracted from polyphonic music signals. *2012 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 81–84. https://ieeexplore.ieee.org/abstract/document/6287822/

Schedl, M., Gómez, E., & Urbano, J. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends® in Information Retrieval*, *8*(2–3), 127–261.

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., & Skerrv-Ryan, R. (2018). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4779–4783. https://ieeexplore.ieee.org/abstract/document/8461368/

Shin, S., Yun, H., Jang, W., & Park, H. (2019). Extraction of acoustic features based on auditory spike code and its application to music genre classification. *IET Signal Processing*, *13*(2), 230–234. https://doi.org/10.1049/iet-spr.2018.5158

Smith, S. M., & Williams, G. N. (1997). A visualization of music. *Proceedings. Visualization'97 (Cat. No. 97CB36155)*, 499–503. https://ieeexplore.ieee.org/abstract/document/663931/

Steinmetz, C. J., & Reiss, J. D. (2021). *WaveBeat: End-to-end beat and downbeat tracking in the time domain* (arXiv:2110.01436). arXiv. http://arxiv.org/abs/2110.01436

Swaminathan, S., & Schellenberg, E. G. (2015). Current Emotion Research in Music Psychology. *Emotion Review*, *7*(2), 189–197. https://doi.org/10.1177/1754073914558282

Thaut, M. H., Trimarchi, P. D., & Parsons, L. M. (2014). Human brain basis of musical rhythm perception: Common and distinct neural substrates for meter, tempo, and pattern. *Brain Sciences*, *4*(2), 428–452.

Town, S. M., & Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00088

Van Der Zwaag, M. D., Westerink, J. H. D. M., & Van Den Broek, E. L. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae*, *15*(2), 250–269. https://doi.org/10.1177/1029864911403364

Virtala, P., Huotilainen, M., Partanen, E., Fellman, V., & Tervaniemi, M. (2013). Newborn infants' auditory system is sensitive to Western music chord categories. *Frontiers in Psychology*, *4*, 47528.

Virtala, P., Huotilainen, M., Partanen, E., & Tervaniemi, M. (2014). Musicianship facilitates the processing of Western music chords—An ERP and behavioral study. *Neuropsychologia*, *61*, 247–258.

Wang, Y., Salamon, J., Cartwright, M., Bryan, N. J., & Bello, J. P. (2020). *Few-Shot Drum Transcription in Polyphonic Music* (arXiv:2008.02791). arXiv. http://arxiv.org/abs/2008.02791

Wu, Y.-C., Hayashi, T., Tobing, P. L., Kobayashi, K., & Toda, T. (2021). Quasi-Periodic WaveNet: An Autoregressive Raw Waveform Generative Model With Pitch-Dependent Dilated Convolution Neural Network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 1134–1148. https://doi.org/10.1109/TASLP.2021.3061245

Yu, S., Sun, X., Yu, Y., & Li, W. (2021). Frequency-temporal attention network for singing melody extraction. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 251–255. https://ieeexplore.ieee.org/abstract/document/9413444/

Yu, S., Yu, Y., Sun, X., & Li, W. (2023a). A neural harmonic-aware network with gated attentive fusion for singing melody extraction. *Neurocomputing*, *521*, 160–171.

Yu, S., Yu, Y., Sun, X., & Li, W. (2023b). A neural harmonic-aware network with gated attentive fusion for singing melody extraction. *Neurocomputing*, *521*, 160–171. https://doi.org/10.1016/j.neucom.2022.11.086

Zamm, A., Schlaug, G., Eagleman, D. M., & Loui, P. (2013). Pathways to seeing music: Enhanced structural connectivity in colored-music synesthesia. *Neuroimage*, *74*, 359–366.

Zatorre, R. J., & Baum, S. R. (2012). *Musical melody and speech intonation: Singing a different tune*. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001372

Zhang, J. (2022). Music Data Feature Analysis and Extraction Algorithm Based on Music Melody Contour. *Mobile Information Systems*, *2022*. https://www.hindawi.com/journals/misy/2022/8030569/

Zhu, Y. (2022). Recognition Method of Matching Error between Dance Action and Music Beat Based on Data Mining. *Security and Communication Networks*, *2022*. https://www.hindawi.com/journals/scn/2022/8176863/