ORIGINAL ARTICLE

EmoteControl: an interactive system for real-time control of emotional expression in music

Annaliese Micallef Grimaud [1] · Tuomas Eerola [1]

Abstract

Several computer systems have been designed for music emotion research that aim to identify how different structural or expressive cues of music influence the emotions conveyed by the music. However, most systems either operate offline by pre-rendering different variations of the music or operate in real-time but focus mostly on structural cues. We present a new interactive system called *EmoteControl*, which allows users to make changes to both structural and expressive cues (tempo, pitch, dynamics, articulation, brightness, and mode) of music in real-time. The purpose is to allow scholars to probe a variety of cues of emotional expression from non-expert participants who are unable to articulate or perform their expression of music in other ways. The benefits of the interactive system are particularly important in this topic as it offers a massive parameter space of emotion cues and levels for each emotion which is challenging to exhaustively explore without a dynamic system. A brief overview of previous work is given, followed by a detailed explanation of *EmoteControl*'s interface design and structure. A portable version of the system is also described, and specifications for the music inputted in the system are outlined. Several use-cases of the interface are discussed, and a formal interface evaluation study is reported. Results suggested that the elements controlling the cues were easy to use and understood by the users. The majority of users were satisfied with the way the system allowed them to express different emotions in music and found it a useful tool for research.

Keywords Music · Emotion · Cues · Interactive · Interface · Design

1 Introduction

This paper will focus on music and emotion research, in particular, musical cues and their effect on emotional expression research. Given the large possible feature range of cues related to emotional expression, understanding how these cues operate through traditional experiments with feature manipulations is severely limited. However, in an interactive paradigm where the user is allowed to control the cue levels in real-time, this constraint is largely absent. Such interface is challenging to design and implement in a way that would satisfy users. In this paper, we propose a new real-time interactive system called *EmoteControl* which taps into a direct user experience and allows users

to control emotional expression in music via a selection of musical cues. The main contribution of *EmoteControl* is that it is a tool that can be utilised by non-experts, allowing users without any musical skill to be able to change cues of music to convey different emotions in real-time.

This paper will be structured as follows. First, an overview of previous research in musical cues and emotions will be offered, ranging from the traditional experiment methodologies to the computer systems designed for this research field, and highlighting some shortcomings in previous research that we aim to redress with our system. *EmoteControl* is then introduced in the third section, giving an account of the system design and how the interface works. Specifications of the source material that can be utilised with the interface will then be covered, and possible limitations of the system will be described. The fourth section will describe two interactive use-cases of the system, as well as a formal evaluation to test the system's usability with non-musicians, drawing on Human-Computer Interaction principles [4, 5]. Finally, the paper will conclude with a discussion about the implications of *EmoteControl* as well as potential future research.

✉ Annaliese Micallef Grimaud
annaliese.micallef-grimaud@durham.ac.uk

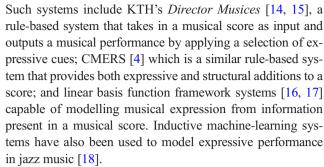1 Music Department, Durham University, Durham, UK

2 Related work

2.1 Musical cues for expressions in emotions in music

Previous literature suggests that listeners can recognise emotions expressed by music due to the composer's and performer's use of musical cues in encoding emotions in music. In return, the listener uses these same cues to successfully decode and recognise the emotion being expressed by the music [1]. Musical cues can be loosely divided into two categories: structural and expressive cues. Structural cues are the aspects of music which relate to the composed score of the music, such as tempo and mode, while expressive cues refer to the features used by performers when playing a piece of music, such as timbre and articulation [2, 3]. Although this distinction has been made between the two categories, which features belong in the structural and expressive categories are still open to interpretation, as some features such as dynamics can be modified by both composers and performers. Thus, it is not always possible to completely separate the performance features from the composition features [4]. Either way, both structural and expressive cues are responsible for conveying emotion [5] and are known to operate together in an additive or interactive fashion [6]; thus, a combination of both structural and expressive cues should be investigated together. In this paper, dynamics, mode, tempo, and pitch will be referred to as structural cues, whilst brightness and articulation will be interpreted as expressive cues.

Throughout the years, empirical studies on emotional expression in relation to musical cues have been approached with different methodologies. Hevner introduced systematic manipulation of musical cues, where structural cues such as mode [7], tempo, or pitch [8] were varied on different levels (e.g. rising or falling melodic contour, or key from major to minor) in short pieces of tonal music, resulting in the creation of distinct versions of the same musical samples. The emotion conveyed in each variation was assessed by participants listening to the different music versions and rating which emotion(s) they perceived, which assisted in identifying how the different cues affected the communicated emotion. Thompson and Robitaille [9] also inspected the structural cues of the music, by identifying similarities and differences in the structure of different melodies. In contrast to this, other studies investigated the effect of expressive cues of the music by instructing musicians to perform music excerpts in different ways to convey various emotions and analyse how listeners use the expressive cues to decode the communicated emotions [1, 10, 11].

Technological advancements have enabled new approaches to music emotion research, as computer systems have been specifically created for modelling musical cues to allow control over the emotions expressed by music [12, 13].

Such systems include KTH's *Director Musices* [14, 15], a rule-based system that takes in a musical score as input and outputs a musical performance by applying a selection of expressive cues; CMERS [4] which is a similar rule-based system that provides both expressive and structural additions to a score; and linear basis function framework systems [16, 17] capable of modelling musical expression from information present in a musical score. Inductive machine-learning systems have also been used to model expressive performance in jazz music [18].

Although the aforementioned systems allow multiple variations of musical cues, particularly, expressive cues, to be generated, this can only be achieved in non-real-time due to the need to pre-render cues. Systems that would carry more applied potential for musical interaction would be real-time systems which do not require pre-defined and pre-rendered discrete levels of cues but allow for control over a continuous range of each musical cue, as well as multiple cue interactions to be explored. A field of research that concerns real-time computer systems and user's interaction with them is Music Human-Computer Interaction (HCI).

2.2 Music interaction and expression in music

Music HCI (Human-Computer Interaction) looks at the various new approaches, methodologies, and interactive computer systems that are utilised for musical activities and music research. Music HCI focusses on multiple aspects of these methodologies and computer systems, such as their design, evaluation, and the user's interaction with the systems [19].

Interactive systems enable a versatile approach to music performance and research, as they allow for real-time manipulation of music. The advantage of the real-time manipulation of expressive cues is that it affords the exploration of the large parameter space associated with altering expression of music, such as the production of different harmonies through a whole body interaction with the system [20, 21], creation of rhythmic improvisation [22], and pitch [23]. These systems are shifting the focus to a direct user approach, where users can directly interact with the system in real-time to manipulate music and emotional expression. For instance, Robin [24, 25] is an algorithmic composer utilising a rule-based approach to produce new music by changing certain structural musical cues depending on the user's choice of emotion terms or inputted values on a valence-arousal model; EDME [26], an emotion-driven music engine, generates music sequences from existing music fragments varying in pitch, rhythm, silence, loudness, and instrumentation features, depending on emotion terms or valence-arousal values selected by the user; the Emotional Music Synthesis (EMS) system uses a rule-based algorithm which controls seven structural parameters to produce music depending on valence-arousal values chosen by the user [27]. Hoeberechts and Shantz [28] created an Algorithmic Music

Evolution Engine (AMEE) which is capable of changing six musical parameters in real-time to convey ten selected emotions, while Legaspi et al. [29] programmed a Constructive Adaptive User Interface (CAUI) which utilises specific music theory concepts such as chord inversions and tonality as well as cadences and other structural parameters to create personalised new music or alter existing ones depending on the user's preferences.

Although the aforementioned systems allow for real-time user interactions using MIDI representation, important shortcomings with respect to gaining insights into the actual musical cues utilised are identified. First, although these past systems work with multiple musical cues, the cues utilised are mainly structural cues and expressive ones are generally disregarded. This issue arises from the fact that most systems work with a real-time algorithmic composer which would require expressive cues to be pre-defined beforehand. Secondly, although the user is interacting with the systems, the user can only partially control the music created or altered, as the user only has access to emotion terms and valence-arousal values. Therefore, the user might not agree with how the music is being created or altered by the system to convey the desired emotion. Thirdly, not all systems have carried out validation studies that confirm that the music generated by the systems is indeed conveying the desired emotion [24, 29, 30]. Finally, these systems are unfortunately not accessible in any form.

By drawing from both music emotion and HCI research, we can summarise that the current offline rule-based systems are capable of modelling several cues simultaneously, but they do not allow for real-time manipulations of the cues and direct interaction with the user. Also, the potential cue space is enormous since there are seven to fifteen cues linked to musical expression which each may be represented on a continuum (e.g. dynamics, pitch, timbre) or at different levels (e.g. articulation, harmony). Mapping the potential cue space to emotions is too large to be exhaustively studied with static factorial experiments, and such experiments would not have any guarantee of having the appropriate cue levels in the first place. On the other hand, current real-time systems allow for partial user interaction; however, the users do not directly change the musical cues themselves. In order to make progress in emotional expression in music research, focus should be directed to real-time systems which allow for the continuous adjustment of musical cues by users without requiring any musical expertise. In this approach, it is productive to allow for a direct user experience, where participants have direct interactions with the system and can personally change cues of a musical piece to convey different perceived emotions as they desire, in real-time [31–33]. This would allow one to explore the potentially large parameter space consisting of cues and cue levels (e.g. consider a simple example for four cues, tempo, articulation, pitch, and dynamics, each having potentially five cue levels, amounting to 1024 cue-level combinations) to emotions in an

efficient and natural way. This latter property refers to the dilemma of traditional experiments where the scholars need to decide the appropriate cue levels (e.g. three levels of tempi, 60, 100, and 140 BPM) which might not be the optimal values for specific emotions and emotion-cue combinations. A real-time system allows to offer a full range of cues to the user which is by far a more ecological valid and effective way of discovering the plausible cue-emotion combinations. Additionally, these interactive music systems should be formally evaluated to confirm their usability and usefulness within their paradigm [34, 35]. Following this line of argument and with the aim of expanding the research on music and emotions from a non-expert perspective, *EmoteControl*, which is a new real-time interactive music system was designed.

3 Interface design

3.1 System architecture

For the creation of the system, two computer programs were used: Max/MSP and Logic Pro X. Max/MSP is used as the main platform while Logic Pro X works in the background as the rendering engine. Additionally, Vienna Symphonic Library[1] (VSL) is used as a high-quality sound synthesizer operated through the Vienna Instruments sample player in Logic.[2] Functions in a Max/MSP patch allow the alterations of different cues of a musical track as it plays in real-time. Music tracks to be utilised in the system are presented in MIDI format, as this allows for better manipulation of the file information than an audio file would.

A MIDI file is read by a sequencer in Max/MSP, which in turn sends the data through to Logic. The output of the MIDI file is played with a chosen virtual instrument in Logic. A chamber strings sound synthesizer from VSL is utilised as the default virtual instrument in the *EmoteControl* interface. During the playback of the MIDI track, structural and expressive cues of the music can be manipulated in a continuous or discrete manner, and changes can be instantly heard as the MIDI track plays in real-time. The system architecture design can be seen in Fig. 1.
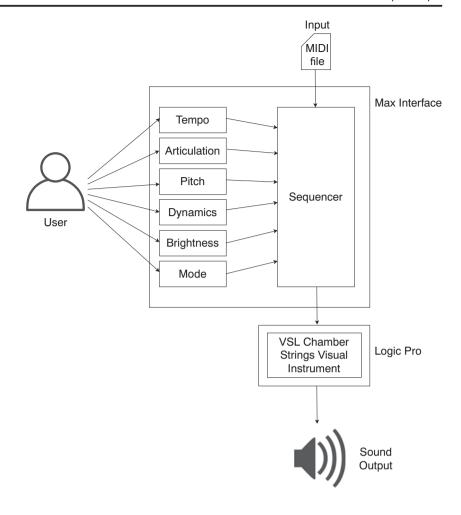
3.2 User interface

EmoteControl (Fig. 2) is targeted at a general audience and aims to allow users to utilise the interface and manipulate cues of a musical piece with no prior music knowledge required. All functions present in the interface are accessed via visual elements manipulated by a computer mouse. The following

---

[1] https://www.vsl.co.at/en
[2] The *EmoteControl* project is available at github.com/annaliesemg/EmoteControl2019
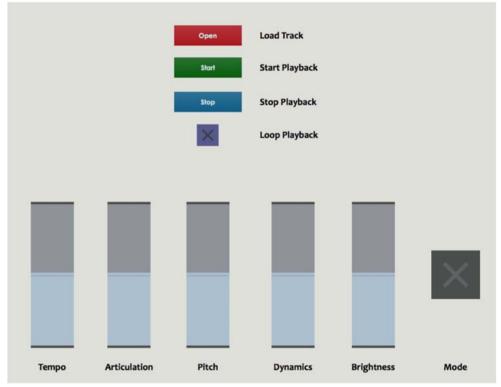
Fig. 1 System architecture of
EmoteControl

Fig. 2 *EmoteControl* user
interface

six cues can be altered via the interface: tempo, articulation, pitch, dynamics, and brightness settings can be changed using the sliders provided, while mode can be switched from major to minor mode and vice-versa with a toggle button.

3.3 Cues

As *EmoteControl* was specifically designed for music and emotion research, the cues available for manipulation were based on previous research, with a selection of frequently studied and effective cues in past research, as well as other less studied cues. Furthermore, the selection of cues features both structural and expressive cues due to their interactive fashion.

Tempo, mode, and dynamics are considered as the most influential structural cues in conveying emotion [36–38]; thus, we have implemented them as possible cues to manipulate to re-affirm and refine the results of past studies, and simultaneously provide a baseline for research carried out with the *EmoteControl* interface. The structural parameter pitch, expressive cues articulation, and brightness have been periodically documented to contribute to emotions expressed by music [39–42], albeit to a lesser degree. Although previous studies have investigated multiple cue combinations [39, 40, 42], most of them were limited to a small number of levels per cue, as different levels had to be pre-rendered. Only a handful of studies utilised live cue manipulation systems to assess the effect of cues on perceived emotions, and these studies were mainly focussed on expressive cues [32, 33, 41].

Thus, the current cue selection (tempo, mode, dynamics, pitch, articulation, and brightness) will allow for a better understanding of how the cue combinations and full range of cue levels affect emotion communication. The six cues available in *EmoteControl* have been implemented in the following fashion.

3.3.1 Tempo

A slider controls the tempo function in Max/MSP which works as a metronome for incoming messages. The function reads the incoming data and sets the speed at which the data is outputted, in beats per minute (bpm). The slider was set with a minimum value of 50 bpm and a maximum value of 160 bpm, to cover a broad tempo range from *lento* to *allegrissimo*. When the slider is moved, the tempo function overrides the initial set tempo, and the MIDI file is played with the new tempo.

3.3.2 Articulation

Three levels of articulation (*legato*, *detaché*, and *staccato)* in the Vienna chamber strings instrument plug-in are controlled via a slider in the Max/MSP patch. The selected articulation methods were put in a sequence from the longest note-duration to shortest, and changes are made as the slider glides from one articulation method to the next. Although the output

of the articulation slider is continuous, the current implementation of the cue is based on real instrument articulation types and therefore offers limited and discrete choices ranging from *legato* to *detaché* to *staccato* levels.

3.3.3 Pitch

A slider in the interface controls the pitch. Pitch shifts are made in terms of semitones, but the implementation of the shift utilises pitch bend function that allows pitch to be shifted in a smooth progression from one semitone to the other. This produces the perception of a continuous pitch shift of the discrete MIDI values of the slider.

3.3.4 Dynamics

The dynamic slider controls the volume feature within VSL. Dynamic changes in a real instrument also affect other acoustical attributes of an instrument, such as the timbre and envelope. In sound libraries such as VSL, sounds are sampled at different levels of loudness to emulate the distinct sound changes produced by the real instruments. On the other hand, changing the dB level of the virtual instrument track will amplify or reduce the overall sound, not taking into consideration the other acoustical properties that are also affected. Thus, controlling the volume within the plug-in produces a more realistic output, although some of the other acoustic parameters (e.g. brightness) will also change. It is worth noting that accenting patterns in the MIDI file (i.e. different velocity values) are retained when dynamic changes are made. The dynamics cue changes the overall level of the whole instrument and therefore retains the difference between the note velocities.

3.3.5 Mode

The mode feature uses the transposer plug-in in Logic to shift the music from its original mode to a new chosen mode and vice-versa. Two types of mode (major and minor) are utilised at the moment, and changes between the two options are made via a toggle on/off button.

3.3.6 Brightness

Brightness refers to the harmonic content of a sound; the more harmonics present in a sound, the brighter it is, while a sound with fewer harmonics will produce a duller sound [43]. A Logic Channel EQ plug-in is utilised for a low-pass filter with a steep slope gradient of 48 dB/Oct, and the Q factor was set to 0.43 to diminish frequency resonance. A slider in the interface controls the cut-off frequency value of the low-pass filter which affects the overall sound output. The low-pass filter has a cut-off frequency range of 305 to 20,000 Hz; the slider in the Max/MSP patch increases/decreases the cut-off frequency depending on the amount of harmonic content desired in the sound output.

3.4 Portable version of *EmoteControl*

In an effort to make the system portable and accessible for field work outside of the lab, a small-scale version of the *EmoteControl* system was configured. This small-scale version of *EmoteControl* also works with Max/MSP and Logic. However, the system does not make use of the extensive VSL library as its sound synthesizer. Instead, one of Logic's own sound synthesizers is used, the EXS24 sampler, making the system Logic-based, albeit more lightweight than a full system with VSL. Substituting VSL with a Logic sound synthesizer makes the system easier to replicate on multiple devices and more cost-effective. Furthermore, it can also be configured on less powerful devices such as small laptops. To limit the hindrance on the ecological validity of the music, a good-quality grand piano was chosen as the default virtual instrument of the portable version of the *EmoteControl* system.

The portable version of *EmoteControl* has a similar user interface (Fig. 3) to the original version and functions the same way, allowing for five cues to be altered: tempo, brightness, mode, pitch, and dynamics. As most of the cue alterations possible in the full version were controlled through VSL, some of the parameters had to be reconfigured. The tempo, brightness, and mode features are external to the sound synthesizer plug-in hence they are not affected by the change in sound synthesizer used. On the other hand, pitch and dynamics settings had to be modified. As articulation changes in the full version of the system were made in VSL, the articulation parameter was omitted from the portable version.

3.4.1 Pitch

The pitch feature in the portable version of the interface allows for incoming MIDI notes to be shifted up or down in semitone steps. Transposition of the music is made stepwise, and the feature allows for a pitch shift up and down 24 semitones from the starting point, a range of 48 semitones in total.

3.4.2 Dynamics

The dynamics cue is controlled by changing the volume (in decibels) of the virtual instrument track in Logic.
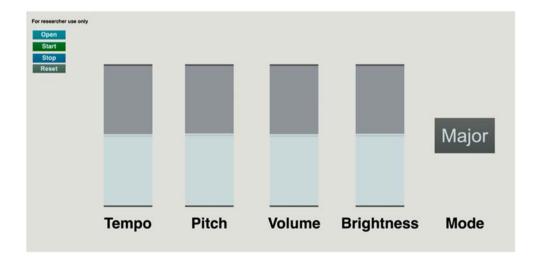
3.5 Source material specifications

There is an amount of flexibility with regard to the MIDI music files which can be inputted in the *EmoteControl* system, although restrictions on the structure of the music compositions also apply, as certain assumptions of the music are held. For the purpose of the research experiment that the system was originally designed for, new musical pieces were composed specifically for the experiment, taking into consideration the necessary assumptions needed for the music to be inputted in the system.

The main assumptions of the music are as follows:

1. Although the interface accepts both MIDI types 0 and 1, music composed should be targeted for one instrument, as the music output will be played using one virtual instrument; thus, multiple parts in the MIDI file will all receive the same manipulations.
2. Music notes should have note durations that allow for different articulation changes of the instrument to be possible (e.g. if a chamber strings synthesizer is utilised as a virtual instrument, then *legato*, *detaché*, and *staccato* articulations should be possible).
3. The pitch range of the music should be compatible with the chosen virtual instrument's register range. As the default virtual instrument of the interface is VSL's chamber strings, the pitch range of B0 to C7 has to be considered.

Fig. 3 Portable version of *EmoteControl* User Interface

4. Music should allow for mode changes (major/minor); therefore, a musical piece should be written in a specific major/minor key in order for the mode changes to be possible.

These assumptions should be taken into consideration and ideally followed to achieve reasonable and natural sounding variation in the cues across a musical piece.

3.6 Interface limitations

As with any system, *EmoteControl* has a number of limitations. A notable constraint is that the interface is based on MIDI, and hence does not support other file formats such as audio files or direct inputs during live performance recordings. The required specifications of the pre-composed music may also be seen as a limitation, as the music compositions to be inputted in the interface have to be purposely chosen or created prior to using the system. Also, the scope of the musical choices does put restrictions in terms of how many MIDI channels are processed, the range of the register across all voices, and non-separation of accompaniment and melodic instruments to mention some of the constraints. Hence, although the interface allows for the live manipulation of musical cues in real-time by the user, the music files have to be pre-made.

Furthermore, although some of the interface's musical cues may be common to most music around the world, making them universal, such as tempo, pitch, and timbre [44–48], it is highly unlikely that *EmoteControl* will fully support non-Western music due to other musical parameters available for manipulation (such as the mode parameter and the exact tuning of the pitches) which vary across cultures. Also, the existing palette of cues may not be optimal for specific genres of Western music, but this interactive framework lends itself to easier prototyping of potential cue-emotion mappings in genres that have so far eluded music and emotion research (e.g. minimalism).

Although the *EmoteControl* system makes use of commercial software, the *EmoteControl* project is readily available and free to the public.[3] Furthermore, the system can be compiled as a standalone application via Max/MSP, which makes the system more accessible.

4 Interface user experience evaluations

As *EmoteControl* is aimed to be utilised by users who do not have any prior music knowledge and investigate their use of musical cues to convey different emotions for music emotion research, the interface was tested and evaluated by multiple

user groups in different contexts, showcasing its use-cases ranging from a science fair to a data collection setting as well as a formal evaluation study utilising HCI methodology. In this section, the various user experiences of both versions of *EmoteControl* will be described, showing how the interface was utilised in various contexts by different users.

4.1 Focus group

First, the interface was presented to a group of seven music emotion researchers in order to gain feedback from researchers who would potentially utilise *EmoteControl* for their experiments, as well as identify any preliminary issues in the interface design prior to its completion. The group session was held at the Music and Science Lab at Durham University which houses the full version of *EmoteControl*. The researchers tried out the interface in a casual group session where they had to alter instrumental music pieces utilising the cues provided and give feedback on potential adjustments and improvements on the interface design, feature labels, parameter ranges, and the system in general. This resulted in alterations such as the redesign of sliders, change in labelling terms, an increase in the tempo parameter range, and utilising a different plug-in to control the brightness feature. Participation in the focus group was on an unpaid voluntary basis.

4.2 *EmoteControl* as a data collection tool

The revised full version of *EmoteControl* was then used as the main tool for data collection in a music emotion research experiment, where both musicians and non-musicians were studied to investigate how the two groups (musicians and non-musicians) altered musical cues to convey different emotions. The study was administered at the Music and Science Lab at Durham University. Full ethical consent was sought and approved prior to the study. Forty-two participants recruited from social media and university communications were given the musical task of manipulating musical pieces via the available cues in the *EmoteControl* interface, in order to convey a particular emotion specified by the researcher. The mean age of participants was 26.17 years (SD = 8.17) and 29 of the participants were female and 12 were male; one participant did not indicate their gender. Twenty-two of the participants were musicians whilst 20 were non-musicians as defined by the Ollen Music Sophistication Index (OMSI) [49]. The participants were first supplied with an information sheet on how the interface works, and a short demonstration was also presented by the researcher. Participants were offered a practice trial before they began the experiment, in order to get them accustomed to the interface and to allow them to ask clarifying questions about the interface.

An open-ended question at the end of the experiment asked participants for any feedback on their experience utilising the interface during the experiment. Feedback revealed that the majority of participants thought the interface design satisfied the requirements for the musical task at hand; participants commented that the interface was clearly labelled and had all the necessary instructions needed for users to utilise it. Six participants added they would have preferred for certain cues to have bigger ranges, such as the tempo range when they were adjusting the tempo of a slow musical piece. Overall, participants felt that they had no issues with understanding the interface, and therefore quickly became accustomed to the interface during the practice trial. All participants were able to successfully complete the task. The whole experiment took approximately 30 min to complete. Participants were remunerated with chocolate.

4.3 Portable *EmoteControl* in an educational context

The portable version of *EmoteControl* was utilised as part of a presentation on music and emotions research at a local science fair known as the Schools' Science Festival, where students interacted with the interface in an educational setting. Nineteen groups consisting of approximately ten school children each, between the ages of 14 and 16 years, attended the University-approved demonstration at individual times. The students were first subjected to a short presentation on how different emotions can be communicated through music. They were then divided into teams of four to five students per team and given the musical task of randomly selecting an emotion term from a choice provided. The aim of the game was to utilise the cues available to alter the pre-composed music supplied and successfully convey the selected emotion term to the rest of the team. A point was given to the team if the other team members correctly guessed which emotion was trying to be conveyed. The game consisted of two or three rounds per team. The team with the most correct guesses in all the rounds would win. Prior to starting the musical task, the students were given instructions and a brief demonstration of how the interface works. The demonstration and musical task had a total duration of approximately 25 min. At the end of the activity, the students were asked if they had any comments about the interface, which led to informal anonymised feedback which was given as a collective. Although students did not necessarily have prior knowledge of the musical cues, the general consensus was that the students found the interface intuitive, and successfully understood what the musical cues do and how to operate it, and for the most part managed to convey the intended emotion correctly to their team members.

4.4 HCI evaluation study of *EmoteControl*

To properly assess and gather in-depth feedback on the design and usability of the interface as a data collection tool for music emotion research by the target users (non-experts), a formal evaluation study was carried out. The field of HCI provides methodologies and software to evaluate devices as well as the user's interaction with the device [50]. Wanderley and Orio [34] drew on methodologies from the field of HCI to provide an evaluation framework for musical interfaces. They propose that the usability of an interface can be assessed by its learnability, explorability, feature controllability, and timing controllability, in a defined context. This can be attained by having target users carry out a simplistic musical task utilising the interface, and afterwards provide feedback. The current evaluation study followed their proposed evaluation method and aimed to assess the usability and controllability of the interface and to provide an overview of how well the interface could be utilised by a general audience.

4.4.1 Method

The study followed a task-based method, where participants were given the task of changing musical pieces via the possible features, in order to make the music convey a specific emotion. Prior to the musical task, the participants were given 2–3 min to get accustomed to the interface. The musical task consisted of three trials featuring different musical pieces. For each piece, the participants altered the music via the six musical features available to make it convey one of the following three emotions: sadness, joy, or anger. The musical task had no time constraint. After the musical task, participants answered both close-ended and open-ended questions pertaining to their experience utilising the interface, usability of the interface, controllability, and range of the features, as well as questions regarding the aesthetic design of the interface. Close-ended questions consisted of 5-point Likert scales, ranging from 'extremely well' to 'not well at all', 'a great deal' to 'none at all', and so on, depending on the question posed. The study took approximately 15 min to complete. Participants were remunerated with chocolate. Full ethical consent was sought and approved prior to testing.

4.4.2 Participants

Twelve participants were recruited via social media and emails. Nine of the participants were female and three were male. Participants ranged in age from 24 to 62 years, with the mean age being 33.67 years. A one-question version of the Ollen Music Sophistication Index (OMSI) [49, 51] was utilised to distinguish between the level of musical expertise. The participants were an equal number of musicians and non-musicians.

5 Results

5.1 Summary of quantitative results

Most of the participants felt they were highly successful in conveying the desired emotion via the interface, with six participants selecting the maximum 'a great deal', five participants selecting the second highest 'a lot', and one participant selecting 'a moderate amount' on the rating scale. All participants indicated that they were satisfied with the usability of the interface to complete the musical task, with nine of the twelve participants selecting the maximum 'extremely satisfied', and the remaining three participants selecting the second highest 'somewhat satisfied' on the rating scale. Nine participants thought that the sliders and toggle button worked 'extremely well' to alter the features, and the remaining three thought they worked 'very well'. All participants were satisfied with the ranges of the features available, with eight participants selecting the maximum 'extremely satisfied', and the remaining four participants selecting the second highest 'somewhat satisfied' on the rating scale. All participants agreed that the interface is extremely easy to get accustomed to, and the terms utilised as labels for the different features were clear, with nine participants selecting the highest 'extremely clear', and the remaining three participants, who were all non-musicians, selecting the second highest 'somewhat

clear' on the rating scale. Figure 4 gives an overview of the participants' ratings for the aforementioned results.

5.2 Summary of qualitative results

Participants were asked open-ended questions on what they liked about the interface, what they disliked, what changes they would make to the interface, and what additional features they would like to see in the interface.

Seven participants mentioned that they appreciated the usability of the interface; five participants directly stated how 'easy to use' the interface is, and two participants said that the interface is 'very intuitive'. Half of the participants stated that they liked the variety of musical cues available, as well as the clear design of the interface. Other comments mentioned they liked the good sound quality of the system and the ability to instantly hear cue changes made to the music as it plays. When asked about what they did not like in the interface, half of the participants commented they had no dislikes, while two participants mentioned the continuous pitch-frequency shift emitted by the pitch manipulation until it stabilises on a new pitch. Individuals noted that the sliders have 'slightly bleak colours', the mode feature is the least intuitive feature for a non-musician, 'the range of the highest and lowest can be more extreme', and that there should be clearer visual
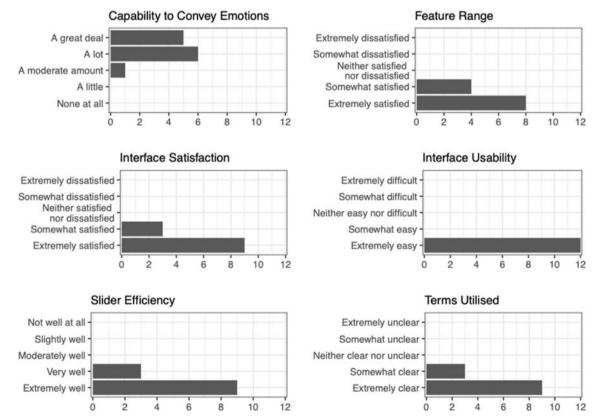

Fig. 4 Participants' responses to the quantitative part of the evaluation study

distinctions between each feature's range, potentially by having colour-changing sliders.

When asked about possible changes the participants would make to the interface, the two most prominent answers mentioned adding terms for the features with definitions to aid non-musicians (25% of comments) and making the interface more colourful (25% of comments). Two participants mentioned expanding the mode feature to include more than the two current levels (major/minor), such as adding an atonal option. Participants were also asked about the interface design, by providing them with the original layout as well as another possible interface layout featuring dials rather than sliders. The alternative design layout presented to the participants was chosen 16.7% of the time, suggesting that the original interface layout was the preferred design, being chosen 83.3% of the time.

The additional features suggested were the option of choosing different instruments (25% of comments), using colours for the musical cues (25% of comments), and adding a rhythm parameter (16.7% of comments). Four participants did not suggest an additional feature. All participants agreed that they would utilise the interface again (100% of participants selecting the maximum of 'extremely likely' on the rating scale) to complete a similar musical task.

Participants were also asked if they thought the interface could be adapted for non-research contexts. All participants agreed that the interface could be utilised outside of a research context, with the two main situations being an educational game for children (mentioned in 58.3% of comments) which would help 'children to get enthusiastic about music', and music therapy, in particular when working with special needs or non-verbal individuals (being mentioned 41.7% of the time). Participants suggested that the interface can be adapted for non-research contexts by installing it on mobile phones and tablets, making it more colourful, and providing definitions for the musical cues and a user guide.

This study was set up to evaluate the interface on its design and usability. The participants' responses indicated that the interface allowed them to successfully carry out the musical task and that they would utilise the interface again for a similar task. This affirms that the aim of *EmoteControl*, which is allowing users to change features of the music to convey different emotions without the need of any musical skills, is possible to achieve. Therefore, the *EmoteControl* interface has potential to be utilised for musical cues and emotion research, which was the primary aim of the project. Furthermore, this study got valuable information from users on what could be the next feature in the interface, such as the possibility of changing the instruments playing, or adding colour to the sliders for a better aesthetic. An interesting outcome was the suggestions for potential adaptation of the interface, where participants strongly agreed on

EmoteControl's ability to be utilised in an education setting, as well as an aid for therapeutic methods.

6 Discussion and conclusions

This paper presented a new interactive system for real-time manipulation of both structural and expressive cues in music. Most importantly, the *EmoteControl* system allows for a direct user experience approach, where users are able to directly interact with the music and change its properties, having potential uses as a tool in musical cues and emotion research, in music composition, and also performance. Furthermore, *EmoteControl* is designed to be utilised by a general population, who do not have significant music knowledge to be able to interact with the interface. To increase accessibility, a portable version of *EmoteControl* has also been designed to be utilised outside of lab settings.

Similar to other systems, certain specifications have to be considered for the system to work at its optimal condition. In *EmoteControl*'s case, the MIDI files to be utilised in the interface have to be compatible with the requirements of the system, mainly concerning the range and the duration of the pitches. However, in the unlikely circumstance that this is not possible, certain settings of the interface can be remapped to accommodate the MIDI file used as input. Potential changes to the interface include utilising a different virtual instrument as output, rather than the default chamber strings virtual instrument. Any virtual instrument from the sound library used in Logic can be selected as the sound output. However, most of the cues available in the *EmoteControl* interface are dependent on the selected virtual instrument; thus, a change in virtual instrument will also affect which cues will be available for manipulation. Therefore, although the aforementioned alterations are possible, the ideal approach would be to utilise music that compliments the system, rather than the other way around.

Possible interface changes would be increasing or decreasing the number of cue levels, such as having more articulation methods to choose from (e.g. adding a *pizzicato* option as well if utilising a string virtual instrument). The range of parameters, such as of tempo and brightness, can be amended as needed. The mode feature can be altered by choosing different mode scales or expanded by adding multiple mode scales to choose from. For instance, the portable *EmoteControl* system has been used for cross-cultural research in Pakistan, where three different scales were being investigated: major, minor, and whole-tone scales. To make this possible, the mode parameter was adapted to incorporate all three aforementioned scales. Another possible alteration would be to add other cues into the system, which would allow for non-Western music to be more compatible with the system, such as rhythm (e.g. event density, expressive

microtiming) or timbre (e.g. shape of the envelope, spectral distribution or onset type) [44, 52] and potentially approaching cues such as mode in an adaptive fashion to reflect the organisation of scales in non-Western cultures. It might also be feasible to allow users to deselect the cues that they feel are irrelevant for the expression of emotions.

Following participants' feedback from the evaluation study, potential changes to the interface design could be the use of different terms as parameter labels to provide a clearer definition of what the parameter does. Given the response of participants about the mode parameter being the least intuitive feature for non-musicians, an example of this could be changing the 'mode' label term to something simpler such as 'major/minor swap' or 'change the pitch alphabet in the music'. Introducing more colour to the interface could also be a possibility, especially as it was frequently mentioned from participants during the formal evaluation study.

In this paper *EmoteControl* has been utilised in different contexts, which include utilising the interface as an interactive game for students in an educational setting and a data collection tool for music emotion research. However, the system has other potential uses. Following participants' response from the formal evaluation study as well as the uses of similar systems [30], *EmoteControl* could be utilised as a therapeutic tool, for instance, as an emotional expressive tool for non-verbal patients [53]. As *EmoteControl* is targeted for non-experts and a general population, it could also be utilised to investigate recognition of emotions in developmental context; Saarikallio et al. [54] report how children who were either 3 years old or 5 years old utilised three musical cues (tempo, loudness, and pitch) to convey three emotions (happy, sad, angry) in music but large individual variation exists at these age stages. *EmoteControl* could be utilised in a similar context, with the opportunity to explore more musical cues which are already programmed in the system. Finally, it could be possible to implement another feature which integrates user feedback about the appropriate cues utilised by others to convey different emotions back into the system. By presenting *EmoteControl* as an online multiplayer game, users have to try and guess each other's cue usage and cue values to convey the same emotion, with the gathered cue information fed back into the system. This method is known as a game with a purpose [55], which might aid in exploring how users utilise the large cue space linked to emotional expression. Utilising this approach could have potential due to its suitability for non-experts and simplicity of the task.

Future research could potentially focus on making the system more accessible, such as reprogramming it as a web-based application, or making it available for tablets and mobile devices [56]. Adding more parameters to the interface, such as the ability to switch instruments while the music plays, or changing the rhythm of the music, would be interesting ventures and allow for more cue combinations to be explored.

Another possible amendment would be the ability to input both audio files [57] and MIDI files into the interface, potentially broadening the system's use-cases. Utilising audio files would allow for more flexibility and a greater amount of source material to be available, as it would be possible to use commercial music and other ready-made audio files, rather than having to specifically make the MIDI source files. Furthermore, an audio format would allow for a more detailed analysis of how the cues were used and manipulated to convey specific emotions utilising audio feature extraction tools such as the MIRToolbox [58] and CUEX [59]. Subsequently, the musical cues available for manipulation would have to be re-evaluated as the data in audio format is represented differently than MIDI files [60, 61]. Cues such as articulation apply changes to individual notes in the music, which would not be possible in audio file formats, as they work with acoustic signals rather than music notation. Therefore, a system which allows for both MIDI and audio files would have to differentiate between the cues that can be utilised depending on the input file format [62].

EmoteControl provides a new tool for direct user interaction that can be utilised as a vehicle for continuous musical changes in different contexts, such as in music performances [63], live interactive installations [64, 65], computer games [3], as well as a tool in a therapeutic context. Most importantly, *EmoteControl* can be used as a medium through which music psychology research on cues and emotion can be undertaken, with the capability of adapting it to suit different experiment requirements, allowing for flexibility and further development in the understanding of the communication of perceived emotions through the structural and expressive cues of a musical composition.

Compliance with ethical standards

References

1. Juslin PN (1997) Emotional communication in music performance: a functionalist perspective and some data. Music Percept 14:383–418. https://doi.org/10.2307/40285731

2. Gabrielsson A (2002) Emotion perceived and emotion felt: same or different? Music Sci VO - Spec Issue 2001-2002:123–147

3. Livingstone SR, Brown AR (2005) Dynamic response: real-time adaptation for music emotion. In: 2nd Australasian conference on interactive entertainment. pp 105–111

4. Livingstone SR, Muhlberger R, Brown AR, Thompson WF (2010) Changing musical emotion: a computational rule system for modifying score and performance. Comput Music J 34:41–64. https://doi.org/10.1162/comj.2010.34.1.41

5. Friberg A, Battel GU (2002) Structural communication. In: Parncutt R, McPherson GE (eds) The Science & Psychology of music performance: creative strategies for teaching and learning

6. Gabrielsson A (2008) The relationship between musical structure and perceived expression. In: Hallam S, Cross I, Thaut M (eds) The Oxford handbook of music psychology, 1st edn. Oxford University Press, New York, pp 141–150

7. Hevner K (1935) The affective character of the major and minor modes in music. Am J Psychol 47:103–118

8. Hevner K (1937) The affective value of pitch and tempo in music. Am J Psychol 49:621–630

9. Thompson WF, Robitaille B (1992) Can composers express emotions through music? Empir Stud Arts 10:79–89

10. Schubert E (2004) Modeling perceived emotion with continuous musical features. Music Percept 21:561–585. https://doi.org/10.1525/mp.2004.21.4.561

11. Juslin PN (2000) Cue utilization in communication of emotion in music performance: relating performance to perception. J Exp Psychol Hum Percept Perform 26:1797–1813. https://doi.org/10.1037//0096-1523.26.6.1797

12. Quinto L, Thompson WF, Keating FL (2013) Emotional communication in speech and music: the role of melodic and rhythmic contrasts. Front Psychol 4:1–8. https://doi.org/10.3389/fpsyg.2013.00184

13. Collins N (2010) Introduction to computer music. John Wiley & Sons, West Sussex

14. Friberg A, Colombo V, Frydén L, Sundberg J (2000) Generating musical performances with director musices. Comput Music J 24:23–29. https://doi.org/10.1162/014892600559407

15. Friberg A, Bresin R, Sundberg J (2006) Overview of the KTH rule system for musical performance. Adv Cogn Psychol 2:145–161

16. Grachten M, Widmer G (2012) Linear basis models for prediction and analysis of musical expression. J New Music Res 41:311–322. https://doi.org/10.1080/09298215.2012.731071

17. Cancino Chacón CE (2018) Computational modeling of expressive music performance with linear and non-linear basis function models. Johannes Kepler University, Linz

18. Ramirez R, Hazan A (2005) Modeling expressive performance in jazz. In: Proceedings of 18th international Florida artificial intelligence research society conference. Association for the Advancement of Artificial Intellifence Press, Menlo Park, pp 86–91

19. Holland S, Wilkie K, Mulholland P, Seago A (2014) Open research online. Choice Rev Online 51:51–2973–51–2973

20. Holland S, Wilkie K, Bouwer A et al (2011) Whole body interaction in abstract domains. In: England D (ed) Whole body interaction. Human-Computer Interaction Series. Springer, London, pp 19–34

21. Bouwer A, Holland S, Dalgleish M (2013) Song walker harmony space: embodied interaction design for complex musical skills. In: Holland S, Wilkie K, Mulholland P, Seago A (eds) Music and human-computer interaction. Springer, London, pp 207–221

22. Gifford T (2013) Appropriate and complementary rhythmic improvisation in an interactive music system. In: Holland S, Wilkie K, Mulholland P, Seago A (eds) Music and human-computer interaction. Springer, London, pp 271–286

23. Wong EL, Yuen WYF, Choy CST (2008) Designing Wii controller: a powerful musical instrument in an interactive music performance system. In: MoMM

24. Morreale F, De Angeli A (2016) Collaborating with an autonomous agent to generate affective music. Comput Entertain 14. https://doi.org/10.1145/2967508

25. Morreale F, Masu R, Angeli A De (2013) Robin: an algorithmic composer for interactive scenarios. Sound mMusic Comput Conf 2013, SMC 2013 207–212

26. Lopez AR, Oliveira AP, Cardoso A (2010) Real-time emotion-driven music engine. Proc Int Conf Comput Creat ICCC-10:150–154

27. Wallis I, Ingalls T, Campana E, Goodman J (2011) A rule-based generative music system controlled by desired valence and arousal

28. Hoeberechts M, Shantz J (2009) Real time emotional adaptation in automated composition. Audio Most 1–8

29. Legaspi R, Hashimoto Y, Moriyama K et al (2007) Music compositional intelligence with an affective flavor. Int Conf Intell User Interfaces, Proc IUI:216–224. https://doi.org/10.1145/1216295.1216335

30. Wallis I, Ingalls T, Campana E, Goodman J (2011) A rule-based generative music system controlled by desired valence and arousal. Proc 8th Sound Music Comput Conf SMC 2011

31. Friberg A (2006) pDM: an expressive sequencer with real-time control of the KTH music-performance rules. Comput Music J 30:37–48. https://doi.org/10.1162/014892606776021308

32. Bresin R, Friberg A (2011) Emotion rendering in music : range and characteristic values of seven musical variables. CORTEX 47:1068–1081. https://doi.org/10.1016/j.cortex.2011.05.009

33. Kragness HE, Trainor LJ (2019) Nonmusicians express emotions in musical productions using conventional cues. Music Sci 2:205920431983494. https://doi.org/10.1177/2059204319834943

34. Wanderley M, Orio N (2002) Evaluation of input devices for musical expression: borrowing tools from HCI. Comput Music J 26:62–76. https://doi.org/10.1162/014892602320582981

35. Poepel C (2005) On Interface expressivity: a player-based study. In: Proceedings of the International Conference on New Interfaces for Musical Expression. pp 228–231

36. Dalla Bella S, Peretz I, Rousseau L, Gosselin N (2001) A developmental study of the affective value of tempo and mode in music. Cognition 80:1–10

37. Morreale F, Masu R, Angeli A De, Fava P (2013) The effect of expertise in evaluating emotions in music. In: … on Music & Emotion (ICME3), …

38. Kamenetsky SB, Hill DS, Trehub SE (1997) Effect of tempo and dynamics on the perception of emotion in music. Psychol Music 25:149–160. https://doi.org/10.1177/0305735697252005

39. Juslin PN, Lindström E (2010) Musical expression of emotions: modelling listeners' judgements of composed and performed features. Music Anal 29:334–364. https://doi.org/10.1111/j.1468-2249.2011.00323.x

40. Quinto L, Thompson WF, Taylor A (2014) The contributions of compositional structure and performance expression to the communication of emotion in music. Psychol Music 42:503–524. https://doi.org/10.1177/0305735613482023

41. Saarikallio S, Vuoskoski J, Luck G (2014) Adolescents' expression and perception of emotion in music reflects their broader abilities of emotional communication. Psychol Well Being 4:1–16. https://doi.org/10.1186/s13612-014-0021-8

42. Eerola T, Friberg A, Bresin R (2013) Emotional expression in music : contribution , linearity , and additivity of primary musical cues. Front Psychol 4:1–12. https://doi.org/10.3389/fpsyg.2013.00487

43. Cousineau M, Carcagno S, Demany L, Pressnitzer D (2014) What is a melody? On the relationship between pitch and brightness of timbre. Front Syst Neurosci 7:127

44. Balkwill L-L, Thompson WF (1999) A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. Music Percept An Interdiscip J 17:43–64. https://doi.org/10.2307/40285811

45. Fritz T (2013) The dock-in model of music culture and cross-cultural perception. Music Percept An Interdiscip J 30:511–516

46. Laukka P, Eerola T, Thingujam NS, Yamasaki T, Beller G (2013) Universal and culture-specific factors in the recognition and performance of musical affect expressions. Emotion 13:434–449. https://doi.org/10.1037/a0031388

47. Thompson WF, Matsunaga RIE, Balkwill L-L (2004) Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. Jpn Psychol Res 46:337–349

48. Argstatter H (2016) Perception of basic emotions in music: culture-specific or multicultural? Psychol Music 44:674–690. https://doi.org/10.1177/0305735615589214

49. Ollen JE (2006) A criterion-related validity test of selected indicators of musical sophistication using expert ratings

50. Carroll JM (2002) Introduction: human-computer interaction, the past and the present. In: Carroll JM (ed) Human-computer interaction in the new millenium. ACM press and Addison Wesley, New York, pp xxvii–xxxvii

51. Zhang JD, Schubert E (2019) A single item measure for identifying musician and nonmusician categories based on measures of musical sophistication. Music Percept 36:457–467. https://doi.org/10.1525/MP.2019.36.5.457

52. Midya V, Valla J, Balasubramanian H et al (2019) Cultural differences in the use of acoustic cues for musical emotion experience. PLoS One 14:e0222380

53. Silverman MJ (2008) Nonverbal communication, music therapy, and autism: a review of literature and case example. J Creat Ment Heal 3:3–19. https://doi.org/10.1080/15401380801995068

54. Saarikallio S, Tervaniemi M, Yrtti A, Huotilainen M (2019) Expression of emotion through musical parameters in 3- and 5-year-olds. Music Educ Res 21:596–605. https://doi.org/10.1080/14613808.2019.1670150

55. von Ahn L (2006) Games with a purpose. Computer (Long Beach Calif) 39:92–94

56. Fabiani M, Dubus G, Bresin R (2011) MoodifierLive: interactive and collaborative expressive music performance on mobile devices. In: Proceedings of NIME 11. pp 116–119

57. Bresin R, Friberg A, Dahl S (2001) Toward a new model for sound control. In: COST G-6 conference on digital audio effects. pp 1–5

58. Lartillot O, Toiviainen P, Eerola T (2008) A Matlab toolbox for music information retrieval. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds) Data analysis, machine learning and applications. Springer Berlin Heidelberg, pp 261–268

59. Friberg A, Schoonderwaldt E, Juslin PN (2007) CUEX: an algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals. Acta Acust united with Acust 93:411–420

60. Cataltepe Z, Yaslan Y, Sonmez A (2007) Music genre classification using MIDI and audio features. EURASIP J Adv Signal Process 2007 . https://doi.org/10.1155/2007/36409, 2007, 1, 8

61. Friberg A (2004) A fuzzy analyzer of emotional expression in music performance and body motion. In: Sundberg J, Brunson B (eds) Proceedings of music and music science Stockholm. Stockholm

62. Oliveira A, Cardoso A (2007) Control of affective content in music production

63. Soydan AS (2018) Sonic matter: a real-time interactive music performance system

64. van t Klooster A, Collins N (2014) In a state: live emotion detection and visualisation for music performance. In: Proceedings of the International Conference on New Interfaces for Musical Expression. pp 545–548

65. Rinman M-L, Friberg A, Kjellmo I, et al (2003) EPS - an interactive collaborative game using non-verbal communication. In: Proceedings of SMAC 03, of the Stockholm Music Acoustics Conference. pp 561–563