Computer Generation and Perception Evaluation of Music-Emotion Associations

Mariana Seiça, Ana Rodrigues, F. Amílcar Cardoso, Pedro Martins, and Penousal Machado

CISUC, Department of Informatics Engineering, University of Coimbra
{marianac, anatr, pjmm, amilcar, machado}@dei.uc.pt

Abstract. Music is intertwined with human emotions as an artistic form with expressive qualities. We present a pilot study of music-emotion associations based on a generative system, which produces parameter-based music to represent four emotions: Happiness, Sadness, Calm, and Anger. To study the perceptual relevance of each parameter, we performed a series of user tests where participants explored multiple combinations of musical parameters to reach a representation for each emotion. Results were compared with the ones from previous studies and empirical experiments proposed by other authors, which gave us a starting point to evaluate each association and discover new possible connections. Although most of the associations were confirmed, a few discrepancies were found, such as the user preference for low pitch in Anger over the expected high pitch. These findings provide better insight and validation of the relationship between music and emotions, and thus a starting point to explore novel representations.

Keywords: algorithmic composition, generative music, music-emotion associations, emotion perception

1 Introduction

Representing emotions through a computational artifact dates from Picard's concept of *Affective Computing*, i.e., how can computers express/recognize affect and gain the "ability to "have emotions" [1]. Picard defended that emotions are an important part of human cognition and perception, and thus have a prominent role in assisting people with computational systems. Music is by nature a subjective field, which resonates with the subjectivity of emotions, and may be the reason why emotions have been so largely used to manipulate music computationally. The most used emotion model for music experiments is the dimensional model proposed by Russell, where emotions are distributed in a two-dimensional space and split by the dimensions of arousal and valence [2, 3]. Based on this model, we chose a set of four discrete emotions evenly distributed in the 2D space (see Fig.1) to ensure a validation of the most perceptually-relevant parameters.

Despite the potential of using emotions to algorithmically compose music, research on this subject is usually limited to a small number of parameters such

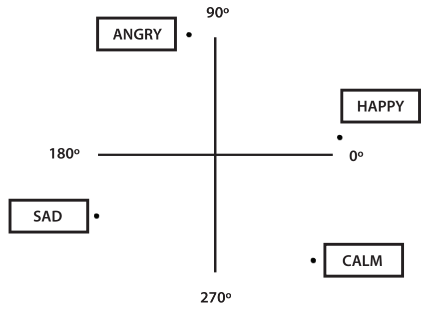

Fig. 1. Choice of emotions based on Russell's Circumplex Model of Affect [2]

as pitch and loudness [4]. In this study, we try to overcome this limitation by developing a computational artifact based on seven musical parameters: harmony, tempo, pitch, melody direction, articulation, melody rhythm, and loudness. To confirm and validate the relationships between these music-emotion associations, we performed a pilot user study, comparing the preferences of each participant to the literature findings on the subject. Our contribution lies in providing a better understanding and support of findings on emotion perception of musical parameters and their respective impact in representing a set of emotions, through the improvement and expansion of a previous generative system [5].

The remainder of this paper is organized as follows. Section 2 comprises an overview of music generation systems, and studies about the association between music and emotions. Section 3 details the improvements made to the previous developed system. Section 4 reports the conducted evaluation, whereas Section 5 provides an analysis and discussion of the results as well as user feedback. Finally, Section 6 draws general conclusions and delineates future work.

2 Related Work

"Music can be used to express emotions more finely than any other language" [1]. But how may these musical characteristics influence the musical forms of expression, and how can computational systems learn this information to produce music based on emotions?

2.1 Music Generation Systems and Music-Emotion Experiments

Algorithmic composition is described as the use of a "formal process to make music with minimal human intervention" [6]. Methodologies used to create automated music range from *stochastic* (Markov chains) and *rule-based*, to *AI* models. A recent work on evolutionary music is Scirea et al.'s *MetaCompose* [7]. With the goal of creating music that can express different mood-states in a dynamic environment, the system generates compositions comprised by "a chord sequence, a melody and an accompaniment" [7], dealing with a set of detailed musical features, such as harmony, melody, pitch, scale, intensity, timbre, and rhythm.

In the line of rule-based models, Livingstone et al.'s computational music-emotion engine [8] presents a set of rules for the score structure and the performative expression over the arousal-valence model, varying musical parameters like the tempo, mode, loudness, articulation, pitch and others to produce, and specifically induce, certain emotional effects in the listener. The term *Generative Music* was popularized by the composer Brian Eno, creating systems that produce ever-changing ambient music through probabilistic rules, such as his first *Generative Music 1* album using SSEYO Koan Software [9], and his last album *Reflection*, available as an infinite piece through an iOS app [10].

David Cope's *Experiments in Musical Intelligence* system is a reference in computer-aided composition, exploring the concept of *recombinance* through the deconstruction of works of classical Western Music, to find common patterns, simulate compositional styles and discover new combinations [11].

2.2 Music and Emotion Associations

Music perception depends on a combination of factors such as an individual's culture, social context, and personality [12]. Although this is a subject commonly discussed and not fully agreed among authors, Juslin has proposed 7 psychological mechanisms [13] through which music may arouse emotions in the listener.

Table 1. Association between a set of music parameters and emotions

	harmony	tempo	melody pitch	melody direction	melody articulation	melody rhythm***	loudness
HAPPINESS	consonant	fast*	high	ascending*	staccato*	dense	loud*
	20%	100%	24%	32%	64%	/	64%
SADNESS	dissonant	slow	low	descending*	legato	sparse	soft*
	20%	100%	36%	16%	56%	/	100%
CALM	consonant **	slow	high*	ascending**	legato**	sparse	soft**
	4%	12%	4%	4%	4%	/	4%
ANGER	dissonant	fast*	high **	ascending	staccato*	dense	loud
	12%	96%	4%	8%	44%	/	100%

* other characteristics for this parameter were found in the literature / ** few experiments found in the literature
*** not found in the literature, based on previous experiments of Seiça et al. [5]

Experiments conducted on music-emotions associations have generally studied the features of loudness, tempo, and consonance/dissonance of harmonies [4, 14]. For example, consonant harmonies have been associated with positive emotions, and loud loudness and fast tempo with emotions of high arousal [14, 15]. A list of the associations found in the literature [4, 14–19] is summarized in Table 1.

Cultural Aspects: Cultural aspects may impact the way we perceive music. For example, in Western music - the one we are working with - melodies played using notes from a major scale tend to be interpreted as happy, while those played with notes from a minor scale usually sound sad [3, 17]. Additionally, the most common chord trajectories of Western tonal music usually begin by establishing a tonal center or base (tonic), and then step away from the stability using more dissonant chords to build tension, to finally return back to the tonic to create relaxation. As a result, the tonic is often the most frequently played note or the note with the longest duration [19].

3 The System: Affective Music Generation

This work is an improvement of a system previously developed by Seiça et al.[5], whose focus was to musically represent emotions retrieved from Twitter. Built as a rule-based system through probabilities, it was guided by two major musical aspects: harmony and melody. The system worked as a communication between three tools: a Processing sketch for analysing the tweets, a Max patcher for the MIDI generation, and an Ableton live set to produce the final sounds.

According to the probabilities defined for each emotion, the melodic line was shaped based on a melodic scale, defined by the harmonic progression to connect to certain emotional contexts. This setting established the set of possible notes for the melody, specifying the type (scale note, chord note or chromatism), the duration of each note (from whole to eighth notes), and the intervals between them, which shaped the melodic motion and leaps. The harmony was defined through predefined harmonic progressions, which combined the affective nature of different chord natures, played in three possible voicings, and their sequence to represent each emotion. The choice of tone quality had a relatively free structure, where timbres and technical features of synthesized sounds were associated with each emotion, resembling the ambient music genre. For further details on the probabilities implemented and the system, we refer the reader to Seiça et al. [5].

3.1 New Parameters and Variations

In this work, we sought to enhance the system by adding new parameters based on the collected studies, and test the relevance of each musical feature in emotion perception. With this purpose, the values for each parameter were simplified.

For the melody, the type of notes was reduced to two sets (Fig. 2 A): the one originally defined for Anger, and the one for Happiness and Sadness. The same was applied to the rhythm (Fig. 2 B): we maintained two sets of different note durations, a denser and a sparser one, to distinguish the emotions with higher energy (Happiness and Anger), and lower (Sadness and Calm). The intervals were reduced to just one set for all emotions (Fig. 2 C), as it was a parameter which we chose not to evaluate, and hence best to have a neutral role. The melody direction (Fig. 2 D) was one of the new parameters, which defines a tendency for creating ascending or descending melodic lines.

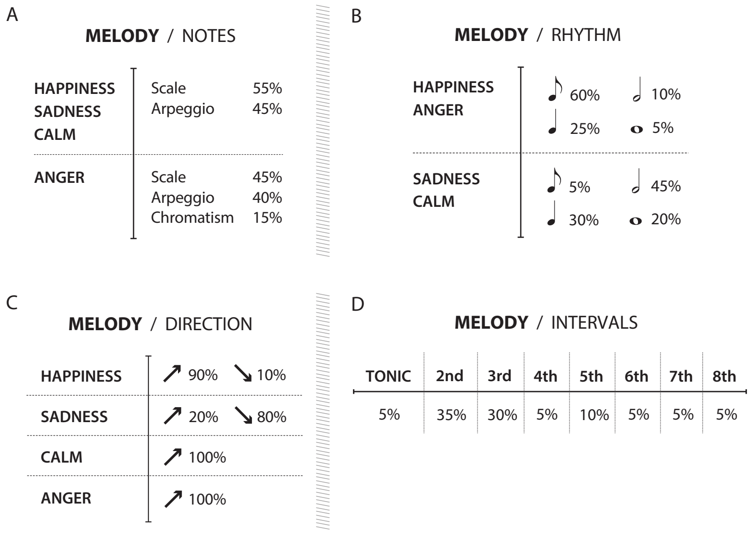

Fig. 2. Probabilities for the type and duration of melody notes, intervals between them and melodic line direction

For the harmony, the progressions were maintained (see examples in Fig. 3 B), and a new parameter was added to control the harmony (Fig. 3 A), playing either a more *consonant* or *dissonant* sound. For the positive emotions (Happiness and Calm), consonance keeps the chords intact, and the dissonance adds notes to the established chords: for example, in a major chord, a minor second, minor third, augmented fourth, augmented fifth or minor seven can be added to the chord structure, all with equal probability. For the negative emotions (Sadness and Anger), dissonance retains the chord structure, and the consonance transforms the dissonant notes of the chords (augmented fourths, augmented fifths and minor fifths) in their consonant counterparts, as perfect fourths and fifths.

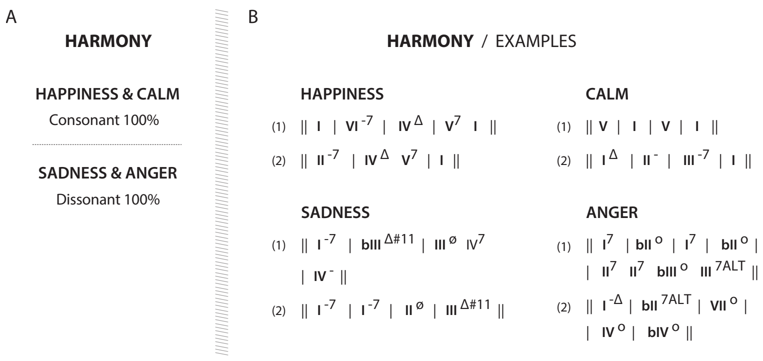

Fig. 3. Examples of harmony progressions and harmony stability for each emotion

The choice for the octave was transformed (Fig. 4 A) to a binary choice of *high* or *low* pitch, which defines the range of possible octaves, with high corresponding to the 4th or 5th, and the low to the 2nd and 3rd. The articulation dynamics was added (Fig. 4 B), which establishes a difference in the harmony and melody dynamics: it can adopt a *legato* style, with each note being played smoothly and connected with each other, or *staccato*, with shorter, detached notes.

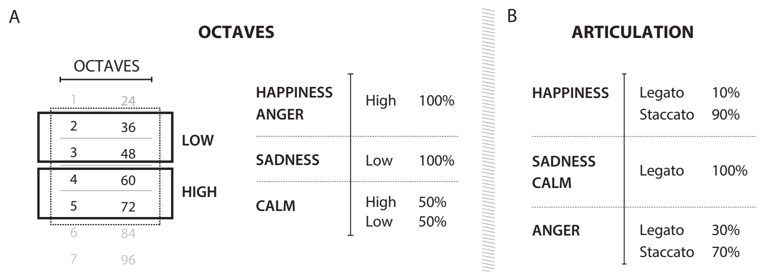

Fig. 4. New octave division, pitch probabilities, and articulation style for each emotion

The control of tempo and loudness, which had already been identified as relevant parameters [5], was implemented as follows. The tempo (Fig. 5 A), measured in BPMs, was defined within three possible range of values: slow (20-75), moderate (76-119) or fast (120-200). Loudness was divided in three MIDI volume levels (Fig. 5 B): soft (44-71), moderate (72-99) and loud (100-127).

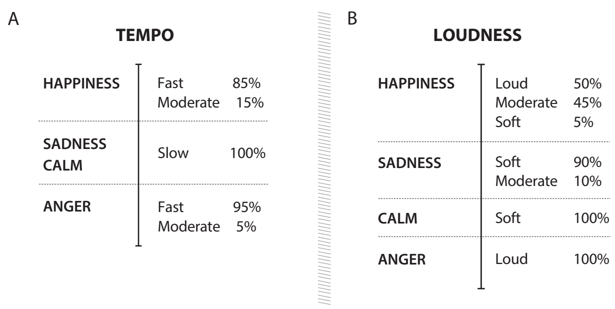

Fig. 5. Probabilities for the tempo and loudness for each emotion

The tone quality was simplified, opting for a piano and a violin, the first to play both the melody and the harmony, and the second for the melody. The choice to reduce the number of timbres to just two instruments, whose sound is familiar and well-recognized, was to balance the tone influence in the emotion association. Therefore, we could focus on evaluating the perceptual weight of the chosen musical parameters with the minimum influence of other characteristics.

4 System Evaluation

We conducted a first set of tests to evaluate the perceptual importance of each musical parameter in representing the four chosen emotions. Seven parameters were tested, according to the number of references in the literature, with the others kept immutable for proper evaluation. The parameters were: harmony, tempo, pitch, melody direction, melody articulation, melody rhythm and loudness. The melody rhythm was an exception we chose to test, despite little mention in the literature review, as it was already implemented in Seiça et al.'s system [5], and we considered it as a relevant feature to evaluate. For each parameter, and based on the literature findings, we established the expected values for each emotion (see Table 1), which would then be compared to the participant's choices.

4.1 Experiment Setup

Twenty participants (12 male and 8 female) took the test. Ages spawned from 22 to 45 years old with an average of 27.8 years and a standard deviation of 4.88. We focused on gathering a balanced set of participants in terms of musical background, distinguishing the ones who have studied music outside the school system and thus have more sensibility to certain musical aspects, from the ones who have not. The tests were performed in person to ensure that the environmental conditions were the same for all the participants.

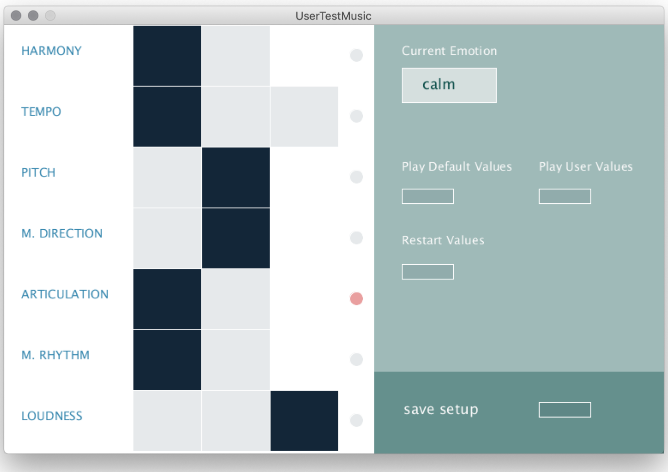

Fig. 6. Interface created for the user tests

We conceived an interface to provide a natural and easy way for participants to interact with our system, which would allow the users to explore the possible combinations between musical parameters (see Fig. 6) in real time, listen to them, and choose the one that most resembled each emotion. The participants had to aurally interpret the impact of the values for each parameter, with their written designation as the only hint. If the participant, for instance, considered a musical parameter to be irrelevant in the emotion representation - either because he/she could not understand the musical variations or the parameter would not influence the representation (in a positive or negative way) - he had the possibility to mark it through a button for it to be deemed as “indifferent/neutral”.

The test would begin with a random combination of musical parameters chosen by the system. This particular choice was to ensure that the participants would not be influenced by the system, and to avoid possible fatigue during the test. The participants also had the possibility, at any point, to listen to the starting point of the system and compare it to their preferences, or even restart the values to the initial set. Once a participant reached the best possible combination, he/she would save the chosen set and evaluate it from 1 to 5 (Likert scale) according to its perception of the emotion representation. In this interval of exchange of feedback there was no music being played, so that the participant could have a moment to refocus and return to a neutral state of mind before the next emotion. This process would repeat through the following order: Happiness, Sadness, Calm and Anger. As the initial set of parameters for each emotion was always random, the order of emotions would not influence its perception: the music could start close or distant from the expected combination of parameters, ensuring a non-biased user’s choice.

5 Analysis of Results

For each emotion we have analyzed: (i) the time each participant took to reach a preferred combination; (ii) the satisfaction/resemblance of his/her combination with the represented emotion; and (iii) the relationship between the participant’s answers to literature findings. We also analyzed results taking into consideration the musical background of each participant.

Regarding **time**, people with no musical background took longer to grasp the musical parameters and the changes caused by each value. This difference was more pronounced in Happiness, with an average of people with musical background taking 3:37 minutes to reach a desired combination, and the ones without taking 5 minutes. Happiness was followed by Calm, with 1:05 minutes distinguishing the participants with and without musical background, Anger with 30 seconds, and Sadness being the most balanced, with just 3 seconds.

The participants **evaluation/resemblance** of each emotion (see Fig. 7) was measured with a Likert scale. According to this assessment, Sadness was considered the emotion with the best musical representation, with 11 participants having chosen the highest classification (5) of Likert scale. The three remaining emotions had their most popular classification in the fourth value. Above all, people with musical background had a tendency to give a higher classification, which can be justified by a more accurate understanding of music fluctuations.

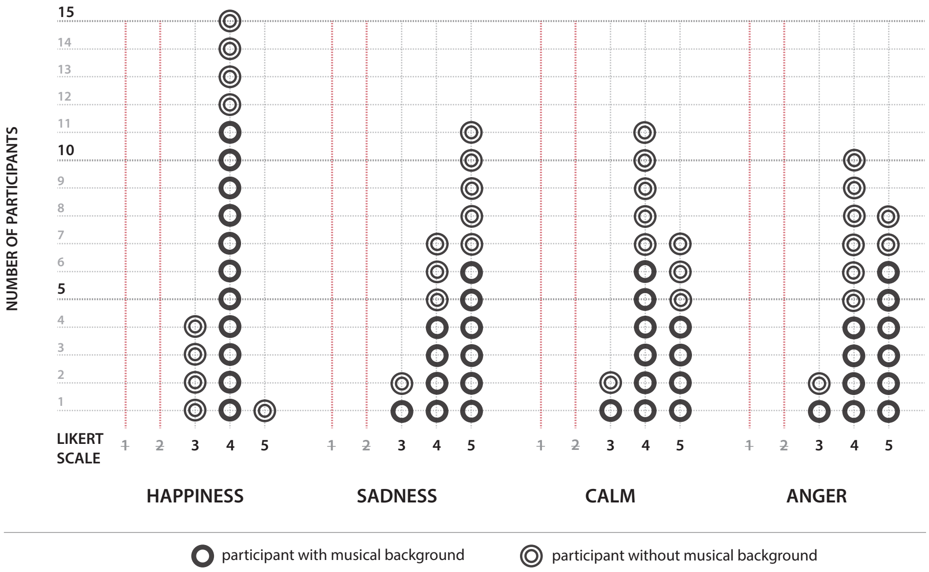

Fig. 7. Likert evaluation for the representation of each emotion

Concerning the **relationship/distance** between participant's preferences and literature findings we performed two types of analysis: (i) distance to the state of art; and (ii) the percentage of correct answers of each music parameter.

The **distance** between the sets of values chosen and the one proposed in the literature is represented in Figure 8. There we can see that despite Happiness having less positive results in the Likert evaluation, had the lowest variation, maintaining the distance between 0 and 3 parameters, which shows a higher correlation with the findings from the literature. Sadness and Anger didn't have a single chosen set equal to the expected, with answers diverging along 6 parameters. Overall, the majority of answers are placed between the distances of 1 and 4, which shows a perceptual tendency and a high correlation with findings from other authors.

As for the **percentage**, we did sum the number of answers that matched the literature findings. Overall, tempo and loudness were the parameters with most answers reflecting these findings, followed by melody articulation, melody rhythm, harmony, melody direction and pitch. One value that stood out was the pitch parameter in Anger, where none of the participants chose the expected value. This may be explained by the scarce number of studies we found in the literature or because it is more perceptually relevant. Happy and Calm were the emotions with more answers matching the literature. Regarding the influence of musical background, there was a fine balance in the answers of participants with and without it, so no major conclusions were drawn in this matter.

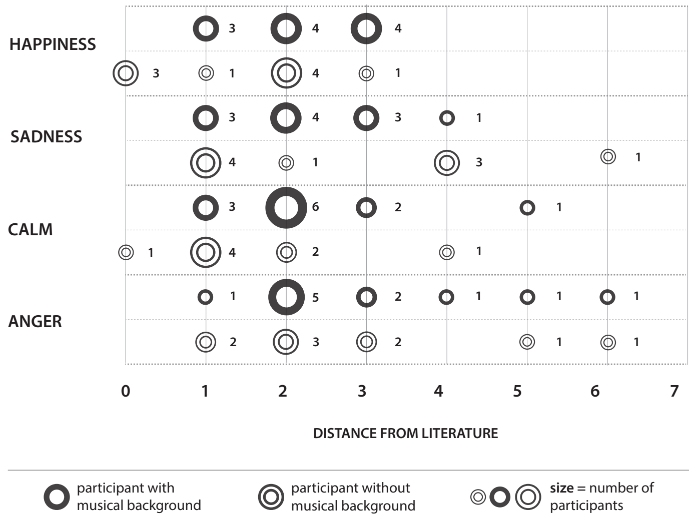

Fig. 8. Distance of the user's choice of parameters to the studies found in the literature. "0" represents the answers matching the literature findings, and "7" no match (all 7 parameters were different)

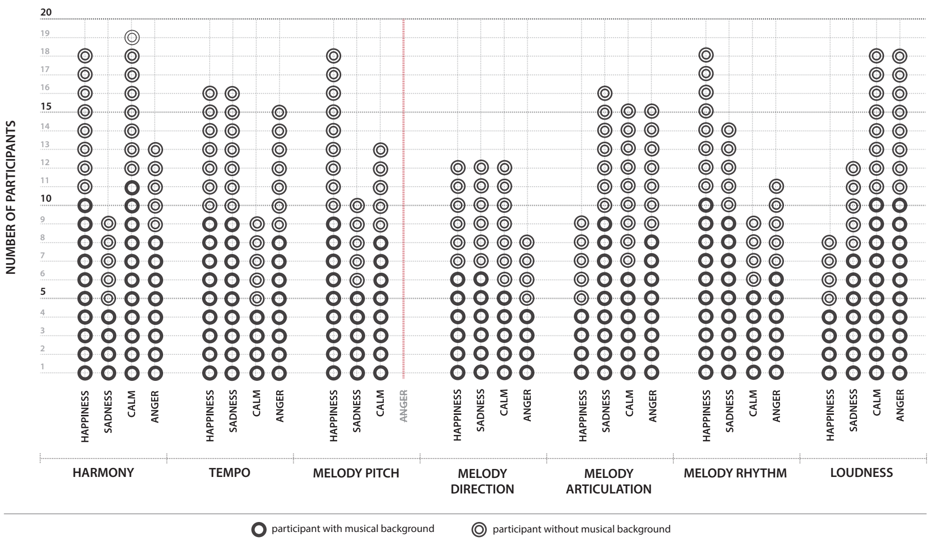

Fig. 9. Distribution of the participants, divided by the musical background, who chose the expected answer concerning each musical parameter

5.1 General Discussion and User Feedback

Overall, the results of the user testing have confirmed a tendency towards the literature. We highlight the difference between participants with and without a musical background in the time they took to perform the test. Because participants with musical background did grasp musical fluctuations more accurately they were faster concluding the test when compared to users with no musical background. However, as suggested by Juslin [12, 13], there are a series of externals factors that may influence the way we perceive music. For example, we observed that most participants performing our test would compare the system's output to previously known songs or contexts - *evaluative conditioning* [13]. They would then explore different combinations of parameters with the goal of finding the combination that would arouse the same kind of emotions. We also reported that at least four participants made a strong association with cinematic scenarios - *visual imagery* [13]. For instance, they reported to imagine a scenario from a movie while listening to the resulting composition of our system.

Concerning the perceptual relevance of musical parameters, melody direction was considered the feature with less impact in the emotion representation, having been reported by seven participants out of twenty. Melody rhythm and harmony were the succeeding parameters, both with four "indifferent/neutral" answers, followed by tempo with two answers, and articulation with one report.

As for user feedback, five participants reported to feel difficulty in recognizing emotions, as their concepts of emotion "relied a lot on the cultural and musical background". Furthermore, these concepts are volatile, and thus its subjectivity, as there is not just one kind of each emotion: Calm can be happier or sadder, Sadness can be more melancholic, anguishing or even nostalgic.

Regarding musical parameters, four participants reported that Happiness should have a faster rhythm and marked pacing, as they considered it to be a key element in a deeper perception of Happiness. Five participants also noted the lack of intermediate values for some parameters, which would allow for more combinations and progressive variations.

6 Conclusion and Future Work

We presented an improved version of an emotional music generative system developed by Seiça et al. [5], with new parameters found in the literature as being relevant on music-emotions association. We used this system to perform an evaluation of these findings through a pilot user-test where most tendencies were confirmed. A few unexpected values were found, such as the low pitch in Anger, which was preferred by all the participants over the expected high pitch.

In future work, we expect to expand the musical features to enrich the musical scenario (e.g. timbre, rhythm) and perform an extended user test, with a larger sample of participants. The statistical analysis will also be detailed with pairwise comparisons to assess the significance of variations and sustained validation.

This pilot study was a first step to confirm general tendencies in emotion representation. These findings can be used to build an audio-visual computational artifact that evolves and generates outputs based on each individual's preferences, exploring the perceptual relevance on the visual domain and how it can be intertwined with the musical domain.

7 Acknowledgements

Mariana Seiça and Ana Rodrigues are funded by Fundação para a Ciência e Tecnologia (FCT), Portugal, under the grants SFRH/BD/138285/2018 and SFRH/BD/139775/2018. We also thank the twenty participants of our study.

References

1. Picard, R.W.: Affective computing. (1997)
2. Russell, J.A.: A circumplex model of affect. Journal of personality and social psychology **39**(6) (1980) 1161
3. Brattico, E., Pearce, M.: The neuroaesthetics of music. Psychology of Aesthetics, Creativity, and the Arts **7**(1) (2013) 48
4. Gabrielsson, A., Lindström, E.: The role of structure in the musical expression of emotions. In: Handbook of Music and Emotion: Theory, Research, Applications. Oxford University Press, New York, NY, US (2010) 367–400
5. Seiça, M., Lopes, R.B., Martins, P., Cardoso, F.A.: Sonifying twitter's emotions through music. In: Music Technology with Swing - 13th International Symposium, CMMR 2017, Matosinhos, Portugal, September 25-28, 2017, Revised Selected Papers. (2017) 586–608
6. Edwards, M.: Algorithmic composition: Computational thinking in music. Communications of the ACM **54**(7) (July 2011) 58–67
7. Scirea, M., Togelius, J., Eklund, P., Risi, S.: Affective evolutionary music composition with metacompose. Genetic Programming and Evolvable Machines **18**(4) (2017) 433–465
8. Livingstone, S.R., Mühlberger, R., Brown, A.R., Loch, A.: Controlling musical emotionality: An affective computational architecture for influencing musical emotions. Digital Creativity **18**(1) (2007) 43–53
9. Software, K.: Generative music) (2018) `https://intermorphic.com/sseyo/koan/generativemusic1/`.
10. Eno, B.: Brian eno reflection (new ambient music) (2017) `https://brian-eno.net/reflection/index.html`.
11. Cope, D.: Experiments in musical intelligence `http://artsites.ucsc.edu/faculty/cope/experiments.htm`.
12. Juslin, P.N., Västfjäll, D.: Emotional responses to music: The need to consider underlying mechanisms. Behavioral and brain sciences **31**(5) (2008) 559–575
13. Juslin, P.N.: From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. Physics of life reviews **10**(3) (2013) 235–266
14. Juslin, P.N., Laukka, P.: Improving emotional communication in music performance through cognitive feedback. Musicae Scientiae **4**(2) (2000) 151–183
15. Scherer, K.R., Oshinsky, J.S.: Cue utilization in emotion attribution from auditory stimuli. Motivation and emotion **1**(4) (1977) 331–346
16. Hevner, K.: Experimental studies of the elements of expression in music. American journal of Psychology **48**(2) (1936) 246–268
17. Balkwill, L.L., Thompson, W.F.: A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. Music perception: an interdisciplinary journal **17**(1) (1999) 43–64
18. Wedin, L.: A multidimensional study of perceptual-emotional qualities in music. Scandinavian journal of psychology **13**(1) (1972) 241–257
19. Cariani, P.: Music perception and cognition (2009)