

Collaborating with an Autonomous Agent to Generate Affective Music

Q1 FABIO MORREALE, Queen Mary University of London
ANTONELLA DE ANGELI, University of Trento

Multidisciplinary research recently has been investigating solutions to offer new experiences of music making to musically untrained users. Our approach proposes to distribute the process of music making between the user and an autonomous agent by encoding this collaboration in the emotional domain. In this framework, users communicate the emotions they wish to express to Robin, the autonomous agent, which interprets this information to generate music with matching affective flavor. Robin is taught a series of basic compositional rules of tonal music, which are used to create original compositions in Western classical-like music. Associations between alterations to musical factors and changes in the communicated emotions are operationalized on the basis of recent outcomes that have emerged from research in the field of psychology of music. At each new bar, a number of stochastic processes determine the values of seven musical factors, whose combinations best match the intended emotion. The ability of Robin to validly communicate emotions was tested in an experimental study ($N = 33$). Results indicated that listeners correctly identified the intended emotions. Robin was employed for the purposes of two interactive artworks, which are also discussed in the article, showing the potential of the algorithm to be employed in interactive installations.

Q2 CCS Concepts: • **Applied computing** ~ **Sound and music computing** • **Human computer interaction** ~ **Auditory feedback** • **Interaction design** ~ **User interface design**

Additional Key Words and Phrases: Algorithmic composition, music and emotions, musical metacreation, interactive installations

ACM Reference Format:

Fabio Morreale and Antonella De Angeli. 2016. Collaborating with an autonomous agent to generate affective music. *Comput. Entertain.* 14, 3, Article 5 (December 2016), 21 pages.
DOI: <http://dx.doi.org/10.1145/2967508>

1. INTRODUCTION

Musical metacreation is a branch of computational creativity that investigates the capability of an autonomous agent to generate creative musical output on its own [Pasquier et al. 2012]. So far, researchers and practitioners working in this area have been mainly focused on producing software that can (1) improvise with performers playing traditional instruments or (2) autonomously compose new scores offline [Eigenfelt et al. 2013]. Our intuition is to employ metacreative software as a support tool to ease music making, thus opening this activity to musically untrained users. Creating musical experiences accessible to anyone is a challenge that has been increasingly tackled by multidisciplinary research in the past couple of decades [Machover 1996]. In this article, we propose an algorithmic composer that allows users to control some aspects of the composition in real time.

Autonomous agents can be defined as algorithmic solutions to create new music with limited or absent human supervision. Autonomous agents can directly map the user

Q3 Author's addresses:

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2016 ACM 1544-3574/2016/12-ART5 \$15.00
DOI: <http://dx.doi.org/10.1145/2967508>

input into musical and sonic features [Franinovic and Salter 2013]. Alternatively, the input can be arbitrarily mapped into combinations of musical parameters and acoustic events using some kind of representation conceived by the artist. Being an arbitrary decision, the mapping is often unclear and many artists are likely to specifically pursue this aspect to provide their users with ambiguous experiences. For instance, Iamascope processes visual information describing the current status of the installation and maps it into specific pitches [Fels and Mase 1999]. Metaphone detects bio-data from the visitors and uses this information to modulate the frequency and the amplitude of predefined tones [Šimbelis et al. 2014]. In both cases, the mapping between sensed input and musical output is unintelligible to the listeners, who might fail to give a meaning to music and to understand how to control it.

Our work has addressed this challenge through design solutions aimed at increasing the transparency of input-output, so that users can intentionally manipulate the melody they are creating. The idea was to reconsider the process of interactive music making as a meaningful collaboration between the human and an autonomous agent, structuring it as an interaction based on emotions, which are available to everybody, intuitive, and naturally connected with music [Morreale et al. 2014]. As a consequence, metacreative software had to be developed, able to autonomously generate a musical composition and systematically convert user input, described in terms of emotions, into musical rules, which are in turn used to direct the composition.

This article presents a new type of generative system based on a certain type of representation of emotions, which are used as an interactive metaphor to allow the user to control the music. Specifically, the user communicates his or her intended emotions to Robin, an algorithmic composer that interprets this information and immediately reconfigures the composition, so that it mirrors the emotions conveyed by the user [Morreale et al. 2013]. Associations between alterations of musical parameters and changes in the communicated emotions were operationalized following research in the psychology of music [Gabrielsson and Lindström 2010; Juslin and Sloboda 2010]. Robin was manually fed a series of rules that are used to generate original music played by virtual instruments. These rules drive a number of stochastic processes that constantly update the values of seven musical factors (i.e., tempo, mode, sound level, pitch register, pitch contour, consonance, and repetitions), whose combinations best match the intended emotion.

This work falls under the domain of the recent research area of algorithmic affective composers (i.e., autonomous systems that generate music with affective flavor). This branch currently counts only a handful of studies [Hoeberechts and Shantz 2009; Legaspi et al. 2007; Livingstone et al. 2010; Oliveira and Cardoso 2010; Wallis et al. 2011], and it is characterized by a number of shortcomings largely ascribable to the early stage of its development. In particular, none of these systems was systematically tested to validate its capability to communicate the intended response in the listener. Also, the quality of the musical output still has large improvement margins.

The contribution of the present study is threefold. First, it presents a new autonomous agent that creates original music by collaborating in real time with the user employing emotions as a medium. Second, it proposes a methodology to evaluate the capability of an algorithmic affective composer to communicate the intended emotions and to test user liking. Third, it presents two interactive systems where the collaboration between Robin and the user is employed to create music with specific emotion character.

This article is structured as follows. Section 2 reviews existing algorithmic affective composers and the related theoretical foundations grounded in the psychology of music and algorithmic composition. Section 3 introduces Robin, detailing the architecture

and the implementation. Section 4 describes the experimental study aimed at testing Robin. Section 5 presents two interactive applications of Robin: The Music Room and The TwitterRadio. The article concludes with reflections about the implications of this autonomous agent and discusses possible future works.

2. RELATED WORK

The automatic generation of musical output with affective flavor is a multidisciplinary subject that relates closely to the psychology of music and algorithmic composition. While the former investigates the human perception to music variations, eventually drawing a mapping between combinations of musical parameters and perceived attributes, the latter studies the musicians' capability of composing musical scores and employs findings to automatically generate new music.

2.1. Psychology of Music

Research in psychology of music has been long investigating the association between variations of musical factors and changes in the emotional expression [Bresin and Friberg 2000; Bresin and Friberg 2011; Fritz et al. 2009; Gabrielsson and Lindström 2010; Hevner 1937; Meyer 2008]. Two main approaches can be adopted for measuring classifying emotions: the categorical and the dimensional approach. The categorical approach postulates that all emotions can be derived from a finite number of monopolar factors of universal basic affects [Ekman 1992]. This approach was adopted by several experimental studies; yet, severe disagreement about the number and the labels of categories is evidenced [Zentner and Eerola 2010]. The dimensional approach, on the other hand, discredits the assumption of independence, postulating that emotions are systematically related to each other and can be described using a limited number of dimensions. The most common dimensional model was proposed by Russell [1980]. It describes emotions as a continuum along two dimensions: valence, which refers to the pleasure versus displeasure affective state, and arousal, which refers to the arousal versus sleep difference. Even though this model is largely adopted in a wide range of research fields, its limitations were acknowledged by the author himself [Russell 1980]. Among other shortcomings, he noted that the affective states in which the two dimensions are convergent (i.e., positive valence and high arousal, and negative valence and low arousal) occur more frequently than the affective states in which they diverge [Russell 1980].

In the psychology of music, both approaches have been widely employed [Juslin and Sloboda 2010], with a predominance of the dimensional approach [Ilie and Thompson 2006; Juslin and Sloboda 2010; Schubert 1999]. A general consensus suggests that the most expressive parameters are tempo and mode, with a slight predominance of tempo¹ [Gagnon and Perez 2003; Gundlach 1935; Juslin 1997; Rigg 1964]. Regarding these findings in the valence/arousal dimensions, tempo has a major impact on arousal and a minor impact on valence, while mode only impacts on valence (Figure 1) [Gagnon and Perez 2003]. Specifically, fast tempo communicates high arousal and, to a lesser extent, positive valence, while slow tempo communicates low arousal and, to a lesser extent, negative valence. Mode influences valence only: major mode generally communicates positive valence and minor mode generally communicates negative valence [Gabrielsson and Lindström 2010]. Interestingly, music played with diverging conditions of mode and tempo (i.e., major mode and slow tempo, or minor mode and fast tempo) seems to communicate similar, neutral levels of valence [Webster and Weir

¹In most cases, *tempo* describes the quantity of notes for the unity of time rather than simply measuring BPMs. This measure is also known as *note density*. For the sake of simplicity, to the term *tempo* also refers to note density in this article.

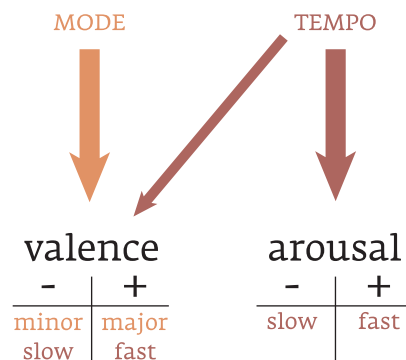


Fig. 1. The double effect of mode and tempo on valence and arousal.

Table I. Mapping Between Musical Structures and the Emotional Dimensions of Valence and Arousal

		Valence	Arousal
Mode	Major	Positive	
	Minor	Negative	
Tempo	Fast	Positive (less influential)	High
	Slow	Negative (less influential)	Low
Sound level	High		High
	Low		Low
Pitch contour	Ascending	Positive	
	Descending	Negative	
Pitch register	High	Positive	
	Low	Negative	
Dissonance		Negative	
Expectations	Fulfilment	Positive	
	Frustration	Negative	

2005]. Nonmusicians, in particular, cannot easily differentiate the valence in musical pieces where valence and arousal diverge [Morreale et al. 2013].

In addition to tempo and mode, other musical factors have a clear influence on the expressiveness of a composition. In particular, we wish to focus on *sound level*, *pitch contour*, *pitch register*, and *dissonance*. This subset of musical factors was selected on the basis of their relevance for communicating emotions [Gabrielsson and Lindström 2010] and their applicability to the architecture of Robin, which mainly operates on structural factors (i.e., those related to the musical score itself), given the objective of algorithmically generating new compositions. The emotional response related to these musical factors is discussed next and summarized in Table I.

- *Sound level*. Sound level is a continuous variable that determines the volume of the musical outcome—that is, the velocity of individual notes. It is directly proportional to the arousal communicated to the listener [Gabrielsson and Lindström 2010].
- *Pitch contour*. The emotional effect of ascending and descending melodic lines has been widely discussed in the literature [Zeiner-Henriksen 2015], but a general consensus on its relevance for emotional expression has not been reached. However, a number of studies have suggested that ascending melodies tend to be associated with positive emotions, while descending melodies are associated with negative emotions [Gabrielsson and Lindström 2010].

- *Pitch register*. High pitch register is associated with positive emotions (but, at times, also fear and anger). Low pitch register is mostly associated with sadness [Gabrielsson and Lindström 2010]. 156
- *Dissonance*. Fritz et al. [2009] suggested that consonance is universally perceived as more positive than dissonance. Moreover, listeners' culture and musical training do not seem to influence the perception of consonance. 157

In addition to these factors, the psychological response of *expectations* is particularly relevant to the emotional response to a musical piece. Related work suggests that the emotional impact of expectation is remarkably complex [Huron 2006]. In general, listener expectations can be either fulfilled or frustrated. Fulfillment/frustration affects the emotional response of the listener [Meyer 2008]. According to this perspective, *resolution* and *repetitions* suggest positive emotions, while *lack of resolution* is indicative of negative emotions. 158

2.2. Algorithmic Composition 159

The algorithmic composition of original music is a creative process combining formal compositional rules with randomness. This combination has been exploited to compose music for centuries. Mozart's *Musicalisches Würfelspiel* ("Musical Dice Game") uses the randomness associated with dice to compose a minuet. Short sections of music are assembled by rolling dice to form a composition with 1.3×10^{29} possible combinations. Given these rules, the musicality of the resulting outcome relied on the coherence of the precomposed sections. Three of the 20th century's finest musicians, John Cage, Iannis Xenakis, and Lejaren Hiller, engaged with a number of compositions that explored stochastic processes for composing music [Schwartz and Godfrey 1993]. In the final decades of the last century, the interest in exploiting randomness in composition resurfaced, partly due to the improved power of computational systems. Computers have been used to develop algorithms capable of generating unpredictable complex structures that are correct from a phraseological perspective [Cope 2005; Jacob 1996; Lewis 1999]. 160

The next three subsections review the most common approaches to algorithmic composition: rule based, learning based, and evolutionary [Todd and Werner 1999]. For a more complete review, refer to Roads and Strawn [1985] and Miranda [2001]. Finally, the last subsection discusses the algorithmic affective composers, an encounter between studies in music perception and algorithmic composition. 161

2.2.1. Rule-Based Approach. The rule-based approach proposes to manually or statistically define a set of compositional rules that provide the system with information on how to compose music autonomously [Boenn et al. 2008; Henz et al. 1996]. These rules drive a number of stochastic processes that generate an original music composition. They can be very basic, as in the previously mentioned *musical dice games* by Mozart, but they can also embody complex harmonization rules [Todd and Werner 1999]. The quality of the music generated with this approach substantially depends on the quality of human intervention, that is, the number of taught rules [Steedman 1984]. As a consequence, metacomposers (those who design the algorithm) need to have a deep knowledge of music theory and a clear sense of their compositional goals. 162

2.2.2. Learning-Based Approach. The learning-based approach proposes to reduce the reliance on human skills. Systems adopting this approach are trained with existing musical excerpts and automatically *learn* compositional rules [Hiller and Isaacson 1957]. Following this approach, Simon et al. [2008] developed MySong, a system that automatically selects chord accompaniments given a vocal track. This study was 163

followed by Songsmith,² a commercial application that empowers users to compose an entire song starting from the vocal track sung by the user. After roughly predicting the notes in the vocal melody, the system selects the sequence of chords that best fits the singing. A music database of 300 excerpts trained a Hidden Markov Model (HMM) that feeds the system with basic statistics related to chord progressions. Another system exploiting the learning-based approach is The Continuator [Pachet 2003], ideated to provide realistic interaction with human players. The algorithm exploits Markov models to react to musical input and can learn and generate any style of music. While this approach reduces the human involvement in the algorithmic composition process, the quality of music is heavily dependent on the training set. Also, this approach is not suitable when there is a need to have direct control of individual musical factors.

2.2.3. Evolutionary Approach. Evolutionary algorithms are stochastic optimization techniques loosely based on the process of evolution by natural selection. Evolutionary algorithms have been used to generate original musical compositions [Mitchell 1996; Miranda 2007]. In most cases, evolutionary compositions attempt to evolve music pieces in the style of a particular composer or genre [Miranda 2007]. In this approach, a population of short, monophonic motifs evolves during the composition. Some other systems also evolve pitch and/or rhythm sequences [Miranda 2007]. In general, the evolutionary approach is particularly effective in producing unpredictable, and at times chaotic, outputs. However, the music might sound unnatural and experimental if compared with rule-based systems, which are generally superior by virtue of the context-sensitive nature of tonal music [Nierhaus 2009]. Furthermore, the evolutionary approach lacks structure in its reasoning and cannot simulate human composers' ability to develop subtle solutions to solve compositional problems such as harmonization [Wiggins et al. 1998].

2.2.4. Algorithmic Affective Compositions. Over the last few years, a handful of studies have attempted to combine theory on music and emotion with algorithmic composition in order to automatically compose expressive music. One of the most interesting examples is AMEE, a patented rule-based algorithm focused on generating adaptive soundtracks [Hoeberechts and Shantz 2009]. The algorithm generates monophonic piano melodies that can be influenced in real time by adjusting the values of 10 emotions with a web applet. The categorical approach was also adopted by Legaspi et al. [2007], who employed an evolutionary approach to composition. Both systems propose interesting methods to adaptive composition, but they employ a categorical approach to emotion classification that fails to address the complexity of the human emotional space (Section 2.1.1).

The dimensional approach can limit this problem, as explained in Livingstone et al. [2010], who follow a rule-based approach that manually collated a set of rules of music theory. The system maps emotions, described along the dimensions of valence and arousal, into structural and performative features. The user interaction is limited to a GUI, where the user can select the desired values of valence and arousal. A similar interface is proposed by Wallis et al. [2011] and by Oliveira and Cardoso [2010] to allow users to interact with the composition. These systems have contributed to defining a novel research topic concerning algorithmic composition of tonal music, allowing users to alter its expressivity in real time. However, a number of significant limitations reduce their practical applicability:

1. The actual capability of the algorithms to communicate correct emotions in the listener has not been validated. The only attempt to determine the extent to which

²Microsoft Corporation. Microsoft Research Songsmith, 2009.

- the listener evaluation of valence and arousal in music corresponded to system parameters was performed by Wallis et al. [2011]. However, the limited number of participants who took part in the study (11), combined with the lack of discussion of the results, undermined the validity of the study.
2. By our estimation, the quality of the music generated by these systems seems to be acceptable only when used for testing the possibilities of an autonomous agent to compose expressive music, rather than being enjoyable by listeners on the basis of its own merits. Again, a formal user study that can disprove this assertion is missing from all of the reviewed literature.
 3. In most cases, the actual interface consists of a simple applet that allows users to select the intensity of discrete emotions, or values of valence and arousal.

This limited utilization, combined with the low quality of the compositions, suggests that these systems are primarily intended as pioneering explorations of a new research field, rather than serving as fully functional systems to be used in interactive contexts. To date, indeed, only Oliveira and Cardoso [2010] have attempted to apply their algorithm to a simple interactive installation, but the audience interaction merely consists of transforming precomposed musical pieces rather than creating music *de novo*.

3. ROBIN: DEVELOPING

Robin was designed to make the experience of musical creation accessible to all users. The system generates a tonal composition in real time while the user interacts with it through control strategies based on basic emotions, described in the valence and arousal dimensions. To ensure consistency with user interaction, the system continuously monitors input changes and adapts the music accordingly, by managing seven musical factors (Section 2.1). As these factors need to be directly accessed and manipulated, a rule-based approach to composition was adopted. This approach allows the designer to manually code the compositional rules and therefore to have full control on the musical factors of interest. As this approach largely relies on human intervention (Section 2.2.1), the quality of the generated music depends on the characteristics and the correctness of the taught rules. To this end, a professional composer was continuously involved at the design and testing stages.

Considering the target user population, a second requirement had to be met: the generated music style had to be understandable even by musically untrained users. For this reason, tonal music was adopted. As opposed to atonal and experimental music, tonal compositions are indeed ubiquitously present in Western culture: even those who lack musical training internalize the grammar of tonality as a result of being exposed to it [Winner 1982]. The process of score generation is grounded upon a number of compositional rules of tonal music driving stochastic processes, which in turn generate *harmony*, *rhythm*, and *melody* (Figure 2). The *harmony* module determines the chord progression following a probabilistic approach. The selected chord is combined with (1) a *rhythmic* pattern that is completed with *pitches* from the scale, thus generating the solo line, and (2) an *accompaniment line selector* that generates an accompaniment line. Finally, the system outputs a stream of MIDI messages that are processed by a Digital Audio Workstation and transformed into music. Robin is currently implemented in SuperCollider.

3.1. Harmony

Traditionally, harmony is examined on the basis of chord progressions and cadences. Following previous works [Nierhaus 2009; Steedman 1984], the transition probabilities between successive chords are defined as Markov processes. Chord transition data can be collected by analyzing existing music, surveying music theory, or following

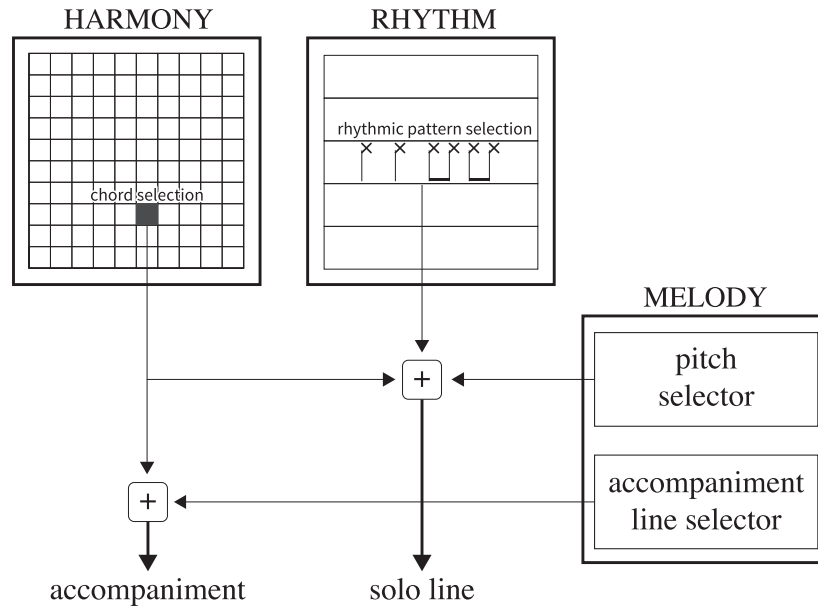


Fig. 2. The architecture of Robin, the algorithmic affective composer.

Table II. Transition Probability Matrix Among the Degrees of the Scale

	I	II	III	IV	V	VI	VII	IV7	V7	II7
I	0	0.05	0.05	0.30	0.20	0.05	0.1	0.05	0.15	0.05
II	0.04	0	0.04	0.04	0.45	0.08	0	0	0.35	0
III	0	0.07	0	0.21	0.07	0.65	0	0	0	0
IV	0.15	0.10	0.05	0	0.35	0.05	0	0	0.30	0
V	0.64	0.05	0.05	0.13	0	0.13	0	0	0	0
VI	0	0.40	0.10	0.10	0	0	0	0	0	0.40
VII	0.8	0	0	0	0	0	0	0	0	0.2
IV7	0	0.30	0	0	0.30	0.30	0	0	0.10	0
V7	0.9	0	0	0.05	0	0.05	0	0	0	0
II7	0	0	0	0	0.5	0	0	0	0.5	0

personal aesthetic principles [Chai and Vercoe 2001]. In our case, a Markov process determines the harmonic progression as a continuous stream of chords. The algorithm starts from a random key and then iteratively processes a Markov matrix to compute the successive chords (Table II). The architecture of the system supports n th order Markov chains. However, for the sake of simplicity, in the current version of the system, chord correlation does not depend on previous states of the system.

The 10×10 matrix contains the transition probabilities among the degrees of the scale. The entries are the seven degrees of the scale as triads in the root position, and three degrees (II, IV, V) set in the VII chord. The transition probabilities are based on the study of harmony presented by Piston [1941]. For each new bar, the system analyzes the transition matrix and selects the degree of the successive bar. The probability for a degree to be selected is directly proportional to the transition value: for instance, with VII the current degree of the scale, the I degree will be selected as the successive chord in 80% of cases on average, whereas the II7 degree will be selected in 20% of

cases. In addition, in order to divide the composition into phrases, every eight³ bars the system forces the harmonic progression to a cadence (i.e., a conclusion of a phrase or a period). Finally, in order to generate compositions with more variability, Robin can switch between different keys performing V and IV modulations.

3.2. Rhythm

For each new bar, a new rhythmic pattern is selected; nearly all rhythm combinations composed of *whole*, *half*, *quarter*, *eight*, and *sixteenth* notes are available. The same combinations of notes in *triplets* are also available. The rhythmic pattern is computed in three steps. First, the time signature of the bar is chosen; second, the values of all the notes played in the bar are selected; third, the selected note values are placed in a particular order. Different time signatures and note values and placement are influenced by two factors, *complexity* and *density*, as follows:

1. *Time signature*. The *complexity* factor determines the time signature: in case of simple rhythms the time signature is duple, triple, or quadruple. By contrast, complex rhythms have irregular time signatures.
2. *Note value selection*. The selection of the values of the notes is influenced by both *complexity* and *density*. In case of simple rhythms, all the notes in the bar have similar values. By contrast, complex rhythms permit notes with very different values to be played in the same bar. In addition, *density* determines the value of the longest note available: very dense rhythms have generally short note values, whereas low-density rhythms are mostly composed of long note values.
3. *Note value placement*. The *complexity* factor also determines the placement of the note values. In case of simple rhythms, notes of the same value are placed one after another, whereas in complex rhythms notes with very different values can be placed nearby.

Figure 3 illustrates a number of rhythmic patterns generated by Robin in 4/4 time signature in the complexity/density dimensions. This technique results in a space of possible solutions with a Gaussian distribution (gray area). Very complex rhythms can happen only when combined with middensity, and very high (low) density necessarily corresponds to very low (high) complexity.

3.3. Melody

In order to generate the melody of the solo line, the rhythmic pattern is filled with suitable pitches. This process happens in three steps:

1. The *pitch selector* receives the rhythmic pattern and the current chord (Figure 4(a)).
2. All the significant notes in the bar are filled with notes of the chord. The notes regarded as significant are those whose duration is an eighth note, or longer, or that are at the first or the last place in the sequence (Figure 4(b)).
3. The remaining spaces are filled with notes of the scale. Starting from the leftmost note, when Robin meets an empty space, it checks the note on the left and it turns it into a higher or a lower pitch, depending on the value of the *pitch contour* (Figure 4(c)). The pseudo-code of the algorithm follows.

The accompaniment line is selected at each new bar. A number of accompaniment lines are available. The accompaniment lines essentially differ in the density of the notes in the arpeggio. Each accompaniment line defines the rhythm of the accompaniment, and the notes of the accompaniment are degrees of the chord.

³Setting the length of a section to eight bars is an arbitrary choice made by the authors. Given the architecture of the system, adopting a different unit or even using a different unit for each section is a feasible option.

5:10

F. Morreale and A. De Angeli

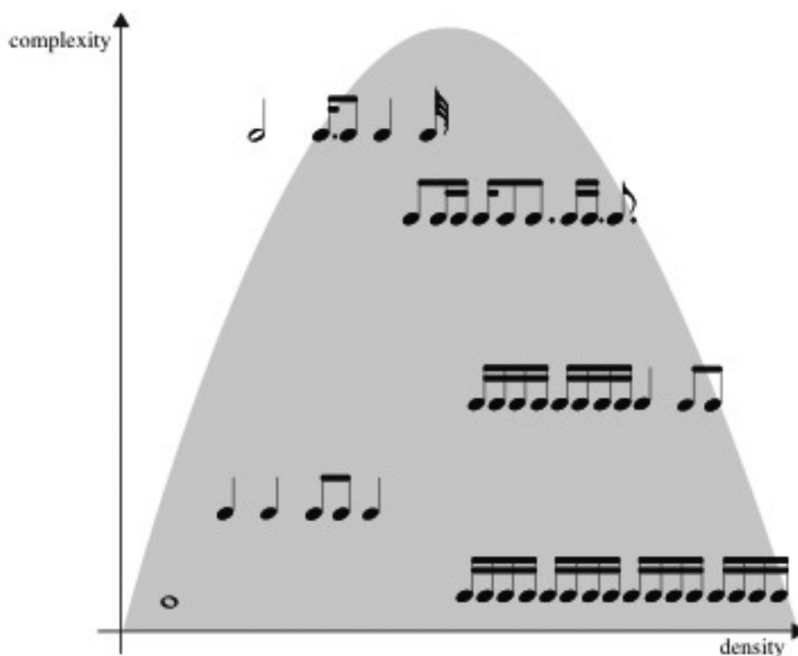


Fig. 3. The gray area represents the space of possible rhythmic solutions.

ALGORITHM 1: Pitch Selection

Input: The rhythmic pattern

Output: The rhythmic pattern filled with pitches
starting from the leftmost note of the bar

repeat

if (*current note value* > *eight note*) **do**

melody(current note) = *random(note from the chord)*

else do

if (*pitch contour* == *ascending*) **do**

melody(current note) = *melody(previous note).next note from the scale*

else do

melody(current note) = *melody(previous note).previous note from the scale*

end

end

until (*number of notes left*) > 0;

3.4. Definition of High-Level Musical Structures

As opposed to similar affective composers such as AMEE [Hoeberechts and Shantz 2009], Robin does not allow the definition of high-level musical structures like phrases and sections. Human composers often make wide use of high-level structures such as phrasing and articulation to create emotional peaks or to develop changes in the character of the composition. However, including such structures in a real-time algorithmic composer is not a viable solution. In order to deal with such structures, the system would need to know the evolution of the piece from the beginning. However, we cannot predict the evolution in advance, as it is controlled by the user. AMEE simulated high-level musical structures by introducing forced abortions in the process of music



Fig. 4. Melody notes selection. (a) The pitch selector receives the rhythmic pattern and the chord. (b) The relevant notes of the melody are filled with notes of the chord. (c) The remaining spaces are filled with notes of the scale to form a descending or ascending melody.

generation [Hoeberechts and Shantz 2009]. However, this solution causes dramatic interruptions, thus reducing both musical coherence and the natural evolution of the composition itself. To this end, the only high-level structural elements manipulated by Robin are repetitions of short themes (which partially simulate choruses and phrases) and cadences (which define phrases).

3.5. Operational Definition of Emotion

Seven musical factors are manipulated to infer changes in the communicated emotions, defined in terms of valence and arousal. These factors are *tempo*, *mode*, *sound level*, *pitch contour*, *pitch register*, *dissonance*, and *expectations* (Table I). This section discusses the operationalization of the alteration of the emotional response of these parameters.

- *Tempo*. Tempo is a continuous variable measured in BPM. Note density is also manipulated by selecting rhythmic patterns and accompaniment lines with appropriate *density* (Section 3.2).
- *Mode*. The change between modes is supported in the *harmony* module, where the chord transition probability matrix is populated with notes based on the selected mode.
- *Sound level*. Sound level changes by manipulating the *velocity* of the MIDI.

Table III. The Value of the Six Factors in the Four Different Conditions of Valence and Arousal

	++	+-	-+	--
Mode	Major	Major	Minor	Minor
Tempo (BPM)	160	80	140	70
Sound level	High	Low	High	Low
P. Contour	Ascending	Ascending	Descending	Descending
P. Register	High	High	Low	Low
Repetitions	Yes	Yes	No	No

- *Pitch contour*. The direction of the melody is determined employing the method described in Section 3.3.
- *Pitch register*. The pitch register center of the compositions generated by Robin ranges from C2 (lowest valence) to C5 (highest valence).
- *Dissonance*. Dissonance is achieved by inserting a number of out-of-scale notes in both melody and harmony.
- *Expectations*. Fulfillment of expectations is operationalized, repeating themes and recurring patterns that the listener quickly comes to recognize as familiar. By contrast, frustration of expectations is operationalized, avoiding repetitions.

4. ROBIN: VALIDATING

This section reports an experimental study aimed at testing the capability of Robin to communicate specific emotions to the listener. For the purposes of this experiment, participants were asked to listen to a number of snippets generated by Robin in different emotional conditions and to self-report the communicated levels of valence and arousal. The experiment could be declared successful if participants correctly identified the intended levels of valence and arousal.

4.1. Procedure

The experimental design was a 2*2 within-subjects design with intended valence (positive vs. negative) and intended arousal (high vs. low). The tested variables were the reported valence and arousal. For each condition, we used Robin to generate five different piano snippets (30 seconds long), for a total of 20 snippets.⁴ A 3-second fade-out effect was added at the end of each snippet. No other processing was made, nor was any generated snippet discarded. Robin manipulated mode, tempo, sound level, pitch contour, pitch register, and expectations⁵ in order to generate music in the four emotional conditions. All the other musical parameters were left constant and high-level structures were not considered at this time.

All factors, except for tempo, influence either valence or arousal. *Tempo*, on the other hand, has a major effect on arousal, but it also influences valence (Table I). This secondary effect is particularly evident for nonmusicians [Morreale et al. 2013]. The double influence of tempo was operationalized as follows:

- Snippets with high arousal were twice as fast as snippets with low arousal.
- Snippets with high valence were 8/7 times faster than the snippets with low valence.

Table III shows the mapping between the six factors and the four conditions of valence/arousal (++ = positive valence/high arousal, +- = positive valence/low arousal, -+ = negative valence/high arousal, -- = negative valence/low arousal). The hypotheses of the study are listed in Table IV.

⁴<http://bit.ly/1HSKjOl>.

⁵Control on dissonance was added to the architecture of the system at a latter stage so it was let out of the experiment. Evaluating the listener's response to changes in dissonance is left to future work.

Table IV. The Expected Values of Valence and Arousal

Intended Emotion	Expected Reported Valence	Expected Reported Arousal
Positive valence – High arousal	+	+
Positive valence – Low arousal	+	–
Negative valence – High arousal	–	+
Negative valence – Low arousal	–	–

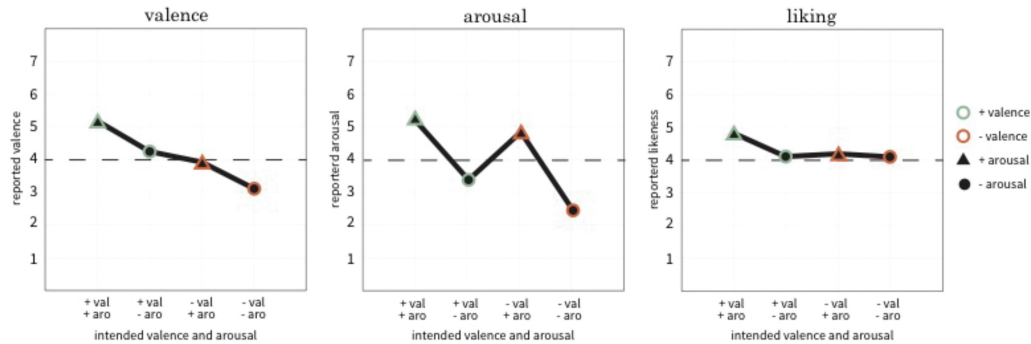


Fig. 5. Graphs describing the averages for the reported valence, arousal, and liking in the four conditions.

Participants were recruited among students and staff of the University of Trento, Italy. A total of 33 participants (11 female, average age 29) took part in the experiment. Sessions ran in a silent room at the Department of Information Engineering and Computer Science. Participants sat in front of a monitor wearing AKG K550 headphones. Before the experiment, participants were given written instructions about the task they had to perform. They were initially presented with four training excerpts in order to become familiar with the interface and the task. Then, the snippets were presented in a random order. In order to measure valence and arousal separately, participants were asked to rate them on two semantic differential items, from 1 (negative or relaxing) to 7 (positive or exciting). In addition, they were asked to indicate, from 1 to 7, how much they liked each snippet (*liking*). To assign the desired value of valence, arousal, and liking, they typed the numbers 1 to 7 on a keyboard when prompted by the interface (e.g., “Please rate 1–7 arousal”). Between each listening, the computer played a sequence of random notes arbitrarily selected from a set of five prerecorded 15-second snippets composed of random notes. Such random sequences are necessary to mask the effects of previously played music [Bharucha and Stoeckig 1987].

4.2. Results

A two-way within-subjects ANOVA was performed on reported valence, arousal, and liking ratings separately. In both cases, intended valence (positive and negative) and arousal (high and low) were the within-subject factors. To disambiguate between the intended valence and arousal (independent variables) and the tested valence and arousal (dependant variables), we will refer to the first couple as *intended* and the second couple as *reported*. We used a p level of .05 for all statistics, and we reported all analyses that reach these levels. The average values of reported valence, arousal, and liking are illustrated in Figure 5.

4.2.1. Reported Valence. The analysis showed significant main effects for intended valence [$F(1,32) = 32.90, p < .001$] and for intended arousal [$F(1,32) = 36.8, p < .001$]. The interaction between the two factors was not significant. As expected, the analysis of the means of the reported valence revealed that + + scored the highest value

Table V. Measured Levels of Reported Valence and Arousal

Intended Emotion	Reported Valence	Reported Arousal
Positive valence – High arousal	+	+
Positive valence – Low arousal	~	-
Negative valence – High arousal	-	+
Negative valence – Low arousal	~	-

(5.21), and -- scored the lowest value (3.22). The double effect of tempo on arousal and valence produced side effects: + - and - + resulted in similar neutral scores (4.26 and 3.91, respectively). These results indicate that the manipulation of valence and arousal contributes to defining the perception of valence, but that the two factors do not intersect. Specifically:

- Regardless of *arousal*, the snippets with positive valence result in more positive values than those with negative valence.
- Regardless of *valence*, the snippets with high arousal result in higher values than those with low arousal.

4.2.2. Reported Arousal. The ANOVA showed significant main effects for intended valence [$F(1,32) = 29.4, p < .001$] and for intended arousal [$F(1,32) = 147.9, p < .001$]. The interaction between the two factors was also significant [$F(1,32) = 12.6, p < .005$]. The analysis of the means of the reported arousal matched our expectations. Snippets composed with high arousal (+ + and - +) scored high values (5.31 and 4.95, respectively), and those with low arousal (- + and - -) scored low values (3.52 and 2.52, respectively). These data suggest that the manipulation of both valence and arousal contributes to defining the perception of arousal, and that their intersection also has an effect, which is evident in the difference between the + - and - - conditions: snippets composed with low arousal communicate higher arousal when combined with positive valence.

4.2.3. Liking. The rating values for each snippet varied between 3.72 and 5.15, with an average of 4.38. The ANOVA revealed that intended arousal was the most significant factor with respect to liking [$F(1,32) = 8.978, p < .01$]. The interaction effect of arousal and valence was also significant [$F(1,32) = 4.735, p < .05$]. The favorite condition was high valence combined with high arousal (mean 4.80, SD .88), while all other conditions produced the same values (4.18, 1.21).

4.3. Discussion

The experiment showed that listeners' emotional responses to the music composed by Robin met our expectations to a significant extent. The reported arousal matched the intended arousal in all conditions. Results on reported valence are more complex. The reported valence matched the intended valence only when the conditions converged. In the case of diverging conditions, the reported valences of - + and + - reported similar, neutral averages (Table V).

This finding can be explained in light of the difficulty experienced by nonmusicians in distinguishing divergent emotional stimulations [Morreale et al. 2013; Webster and Weir 2005]. A possible solution to improve the accuracy of Robin in eliciting correct valence among listeners in these conditions would be to decrease tempo in the - + condition or to increase it in the + - condition. The new values for rebalancing tempo might follow the results of a recent study conducted by Bresin and Frieberg [2011], who suggested that happy performances are usually played almost four times faster than sad performances.



Fig. 6. The Music Room.

5. ROBIN: INTERFACING

This section presents two interactive installations, the Music Room and the Twitter-Radio, in which the generated music results from a collaboration between Robin and the visitors. The contribution to the field of interactive art lies in the employment of an algorithmic composer that is specifically designed to communicate predictable emotions. The input communicates to the system the emotions users want to convey; Robin interprets this information by adapting the values of seven musical parameters to match the desired emotional configuration. The collaboration is mediated by the metaphor of emotions and put into practice through application-specific metaphors. In the Music Room, the user communicates his or her intended emotions via body gestures; in the TwitterRadio, emotions are inferred from textual information describing people's feelings on trending topics.

5.1. The Music Room

The Music Room (Figure 6) is an interactive installation for collaborative music making [Morreale et al. 2014]. The installation was designed to be experienced by couples of visitors, which can direct the emotional character of music by means of their movements. In order to communicate the desired emotions in an intuitive and engaging manner, we adopted the metaphor of intimacy. Distance between people influences valence: the more proximal the visitors are, the more positive the music. The speed of their movements influences arousal: the faster they move, the louder and faster the music is. The process of generating music from user movements involves two steps:

1. Participants' movements are detected using computer vision techniques. The motion of the couples is captured through a downward-looking bird's-eye camera installed

on the ceiling of the room. The detection of the moving subjects has been implemented by applying a standard background subtraction algorithm.

2. The extracted values of average speed and relative distance are communicated to Robin. Following the mapping detailed in Section 3.5, valence and arousal are transformed into combinations of musical factors, which determine the change produced in the generated music. By matching the values of speed and proximity to emotions, Robin adapts the musical flow, as has been previously described.

For the purpose of increasing the diversity and the liking of the composed music, different musical instruments were associated with as many conditions. The *piano* was constantly present in all conditions, a *violin* harmonized the piano voice when couples were particularly close, and a *trombone* harmonized the piano voice when couples were on the opposite sides of the room.⁶ This choice, which was grounded in both personal taste and related work [Eerola et al. 2013; Juslin and Sloboda 2010], was particularly appreciated by the audience of the installation.

Formal evaluations of the installation are reported in previous publications, which describe visitor experiences as collected during three exhibitions of the installation [Morreale and De Angeli 2015]. On each occasion, people were queuing up for a while to try the installation and attendee reviews also seemed to confirm its successful reception. Integrating evidence collected through an array of evaluation techniques disclosed a number of interesting themes. Several visitors reported the feeling of being empowered to create “meaningful” music simply by means of their movements. Others stressed that they had been enabled to have control on music for the very first time in their lives. Furthermore, a quantitative analysis revealed that there was a significant negative correlation between visitors’ musical expertise and engagement, suggesting that nonmusicians had a more creative experience [Morreale and De Angeli 2015].

These results confirmed that the system is capable of offering the audience a unique experience of music making where the control over the composition is shared between the visitor and Robin. Some users argued that they would have preferred to have more control of the music, for instance, by moving their limbs or fingers. We purposely decided to let the user interact on a semantic level only, to ensure a quick engagement with the installation, which might have been hindered by a more complex mapping between user gestures and musical output. However, given the modular architecture of Robin, future editions of the installation could allow users to directly interact with lower-level parameters such as rhythmic complexity and pitch contour.

5.2. The TwitterRadio

The next case study utilizes Robin as a sonification tool for interactive visualization of data. The TwitterRadio offers a novel environment for experiencing user-generated content in an auditory form [Morreale et al. 2014]. The idea is to use music as a means to express data describing public opinions on trending topics. Visitors of the TwitterRadio can browse a list of trending topics and listen to the mood of the world population on those specific topics. The adopted data source is Twitter, which counts over 300 million active users (by May 2015) who constantly share their thoughts and feelings on personal and social issues. The system collects all recent tweets labeled with trending hashtags and retrieves information about their emotions and popularity. These features are then mapped in the musical domain in order to create melodies that match the mood of the tweets. The architecture of the TwitterRadio is composed of three main modules: the user interface, the server, and Robin.

⁶An extract from the Music Room can be viewed at <https://youtu.be/OSVfjVivlw>.



Fig. 7. The TwitterRadio.

1. The user interface resembles a retro-style radio composed of a wooden box, a color display, a knob, and four LED lights (Figure 7). The display shows information about the list of available channels and a red bar indicating the currently playing station. The user can operate the radio by rotating the knob, whose position information is digitalized by an Arduino hidden inside the box. Besides choosing existing trending topics, the user can type his or her favorite hashtag with a wireless keyboard. Finally, the LED lights communicate the status of the system: *playing*, *loading*, or *waiting*. 563 564 565 566 567 568 569 570
2. The server forwards the user requests for a new station to Twitter and gathers all the tweets labeled with that particular hashtag that were posted within the previous 5 hours. The scraped messages are then processed and information about the average tweet mood, the tweet frequency, and the retweet percentage is processed. Tweet mood is computed by means of the MPQA Subjectivity Lexicon [Riloff and Wiebe 2003], which describes the polarity (positive, neutral, negative) of 8,221 English words. The frequency of the tweets is defined as the number of tweets per minute. Retweet percentage refers to the overall amount of retweets. This information is then forwarded to Robin, while data describing the status of the system is displayed through the LED lights. 571 572 573 574 575 576 577 578 579 580
3. Robin collects the information coming from the server and generates music accordingly, diffusing sound through two desktop loudspeakers, which are also hidden inside the box. Tweet mood is mapped into valence and tweet frequency into arousal. Also, when the retweet percentage is above a certain threshold, theme repetition is triggered. Resembling the functionality of traditional radios, when the bar is 581 582 583 584 585

not perfectly aligned with the indicator of a radio channel, the auditory output is buzzing.

The installation was showcased during two academic events held in Trento and at the Art Museum of Rovereto.⁷ A formal evaluation of the experience is currently under study. However, preliminary observations suggested that the audience visibly appreciated both the aesthetic and the functionality of the TwitterRadio and found it particularly entertaining. Furthermore, a number of creative interpretations of the system took place. For instance, some visitors rotated the knob in and out of a channel to rhythmically alternate noise and music, and others tried to create a *song structure* by purposely switching between themes with different moods.

6. DISCUSSION

The work presented in this article provided substantial contributions to the research field intertwining metacreative software with interactive installation. The first contribution is Robin, an algorithmic composer that generates real-time tonal compositions with affective connotations. A possible alternative to the generative approach would have a database of precomposed melodies with different combinations of valence and arousal. However, we maintained that the generative approach would better match our objectives for two main reasons. First, the generative approach permits continuous adaptations of different musical parameters, and therefore nearly an endless number of combinations. The combinatory approach would have required an enormous database of precomposed melodies, which, besides requiring a huge amount of time to be created for each new interaction, would have dramatically increased the size of the software. Second, the generative approach creates a completely new and original composition, thus allowing users to create unique music.

This article also offered a methodology to assess the capability of an algorithmic affective composer to communicate the intended emotions. We validated such capability in an experimental study. This study was of primary importance in that the collaboration between the human and the autonomous agent is encoded with the metaphor of emotions; thus, it was necessary to make sure that the music created by Robin actually stirs among listeners the intended emotional flavor. A systematic validation of the mapping proposed to communicate user meanings is new to the interactive art community. In interactive artwork, mapping strategies have to be defined to transform audience behaviors into musical output. Instead of arbitrarily mapping audience behaviors into musical parameters, we introduced an intermediate layer to mediate users' intentionalities through semantic descriptors translating them into rules that are used by Robin to compose matching music.

Robin put forward other contributions in the field of automatic composition of affective music. In particular, the authors' opinion is that the quality of the music generated by Robin constitutes progress with respect to the music generated by related studies. Moreover, the tunes generated by these systems do not match our personal aesthetic. We believe that this is an important issue that should be taken into primary consideration when discussing works intersecting art and research. This belief echoes the statement of Eigenfeldt et al. [2012], who suggested that metacreative works should reflect the *artistic sentiment* of their designer. Aesthetic and tastes indeed play a crucial role in the evaluation of such systems, which might potentially be flawless from a methodological point of view but still unable to meet the wishes of designers and listeners.

⁷A short video demoing the TwitterRadio can be found at https://youtu.be/GD0a_bNEQCg.

A number of shortcomings, partly ascribable to the infancy of this field, suggest that there are indeed wide margins for future improvement. First and foremost, the evaluation of Robin so far has been limited to validating its actual capability to communicate the intended emotions to the listener. Appreciation of the quality of the music, however, was simply questioned during an exhibition of the Music Room. On that occasion, visitors generally enjoyed the music [Morreale and De Angeli 2015]. In the future, we will aim to set up an experimental study with both experts and nonmusicians to systematically inquire into the quality of the compositions generated by Robin.

Our investigations disclosed that naïve listeners tend to use an emotional vocabulary when describing musical pieces. In the process of simplifying music access to this category of users, then, our first objective was to allow them to have control of the affective flavor of the song. However, we acknowledge that musical grammar is much more complex and can by no means be reduced to an emotional grammar. Future implementation of the system will allow users to interact on other dimensions too.

Currently, the system does not support high-level musical structures; as music progression cannot be predicted in advance, the evolution of the piece is under the control of the user. Should high-level structures be included in the system, sudden input changes from the user would make the transition unnatural. This issue remains open to investigation.

The current implementation of Robin only deals with structural factors to infer a change in the communicated emotions. To enhance communication of the correct emotional flavor, future implementations of the system will include those performative behaviors whose variations define a change in the communicated emotions. Phrasing, for instance, has a direct effect on the communicated emotions: forward phrasing is usually associated with sad and tender performances, whereas reverse phrasing is usually associated with aggressive performances [Bresin and Friberg 2011]. The real-time score generation capability of Robin can be easily combined with existing systems for the automatic modeling of expressive contents of musical scores, such as pDM from Bresin and Friberg [2011], in which performative factors are mapped into emotions.

7. CONCLUSION

So far, musical metacreation systems have been mainly designed for the community of musicians, advancing solutions to autonomously improvise with performers or to autonomously generate new compositions [Eigenfelt et al. 2013]. This article proposed a new direction for musical metacreation by employing computational creativity to provide musically untrained users with experiences of music making. We presented a computational system that distributes the complexity of music making between the user and Robin, an autonomous agent that generates music on its own, allowing the user to interact with the composition on a semantic level. This protocol, which was employed in two interactive artworks, proved particularly efficient in light of the boost of musical creativity experienced by the users.

REFERENCES

- Georg Boenn, Martin Brain, and Marina De Vos. 2008. Automatic composition of melodic and harmonic music by answer set programming. In *Logic Programming*. Springer, Berlin, 2008, 160–174.
- Jamshed J. Bharucha and Keiko Stoeckig. 1987. Priming of chords: Spreading activation or overlapping frequency spectra? *Perception & Psychophysics* 41, 6 (1987), 519–524.
- Roberto Bresin and Anders Friberg. 2000. Emotional coloring of computer-controlled music performances. *Computer Music Journal* 24, 4 (2000), 44–63.
- Roberto Bresin and Anders Friberg. 2011. Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex* 47, 9 (2011), 1068–1081.

- Wei Chai and Barry Vercoe. 2001. Folk music classification using hidden markov models. In *Proceedings of International Conference on Artificial Intelligence*. 6, 6.4.
- David Cope. 2005. *Computer Models of Musical Creativity* (p. xi462). MIT Press, Cambridge.
- Tuomas Eerola, Anders Friberg, and Roberto Bresin. 2013. Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology* 4 (2013).
- Arne Eigenfeldt, Oliver Bown, Philippe Pasquier, and Aengus Martin. 2013. Towards a taxonomy of musical metacreation: Reflections on the first musical metacreation weekend. In *Workshop on Musical Metacreation*. 40–47.
- Arne Eigenfeldt, Adam Burnett, and Philippe Pasquier. 2012. Evaluating musical metacreation in a live performance context. In *Proceedings of the 3rd International Conference on Computational Creativity*.
- Paul Ekman. 2009. An argument for basic emotions. *Cognition & Emotion* 6, 3–4 (1992), 169–200.
- Sydney Fels and Kenji Mase. 1999. Lamascope: A graphical musical instrument. *Computers & Graphics* 23, 2 (1999), 277–286.
- Karmen Franinovic and Christopher Salter. 2013. The experience of sonic interaction. *Sonic Interaction Design* 39 (2013).
- Thomas Fritz et al. 2009. Universal recognition of three basic emotions in music. *Current Biology* 19, 7 (2009), 573–576.
- Alf Gabrielsson and Erik Lindström. 2010. The role of structure in the musical expression of emotions. In *Handbook of Music and Emotion: Theory, Research, Applications*. 367–400.
- Lise Gagnon and Isabelle Peretz. 2003. Mode and tempo relative contributions to “happy-sad” judgements in equitone melodies. *Cognition & Emotion* 17, 1 (2003), 25–40.
- Ralph H. Gundlach. 1935. Factors determining the characterization of musical phrases. *American Journal of Psychology* (1935), 624–643.
- Martin Henz, Stefan Lauer, and Detlev Zimmermann. 1996. COMPOzE-intention-based music composition through constraint programming. In *Proceedings of the 8th IEEE International Conference on Tools with Artificial Intelligence, 1996*. IEEE.
- Kate Hevner. 1935. The affective character of the major and minor modes in music. *American Journal of Psychology* (1935), 103–118.
- Lejaren A. Hiller Jr. and Leonard M. Isaacson. 1957. Musical composition with a high speed digital computer. In *Audio Engineering Society Convention 9*. Audio Engineering Society.
- Maia Hoeberechts and Jeffrey Shantz. 2009. Realtime emotional adaptation in automated composition. *Audio Mostly* (2009), 1–8.
- David B. Huron. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.
- Gabriella Ilie and William Forde Thompson. 2006. A comparison of acoustic cues in music and speech for three dimensions of affect. (2006), 319–330.
- Bruce L. Jacob. 1996. Algorithmic composition as a model of creativity. *Organised Sound* 1, 3 (1996), 157–165.
- Patrik N. Juslin. 1997. Perceived emotional expression in synthesized performances of a short melody: Capturing the listener’s judgment policy. *Musicae scientiae* 1, 2 (1997), 225–256.
- Patrik N. Juslin and John A. Sloboda. 2010. *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press.
- Roberto Legaspi et al. 2007. Music compositional intelligence with an affective flavor. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*. ACM.
- Steven R. Livingstone et al. 2010. Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal* 34, 1 (2010), 41–64.
- George E. Lewis. 1999. Interacting with latter-day musical automata. *Contemporary Music Review* 18, 3 (1999), 99–112.
- Tod Machover. 1996. The brain opera and active music. In *Catálogo Ars Electronica*.
- Leonard B. Meyer. 2008. *Emotion and Meaning in Music*. University of Chicago Press.
- Eduardo Miranda. 2001. *Composing Music with Computers with Cdrom*. Butterworth-Heinemann.
- Eudardo Miranda. 2007. *Evolutionary Computer Music*. Springer, London, 2007.
- Fabio Morreale et al. 2013. The effect of expertise in evaluating emotions in music. In *3rd International Conference on Music & Emotion*.
- Fabio Morreale, Raul Masu, and Antonella De Angeli. 2013. Robin: An algorithmic composer for interactive scenarios. In *Proceedings of Sound and Music Computing Conference*.
- Fabio Morreale et al. 2014. Collaborative creativity: The music room. *Personal and Ubiquitous Computing* 18, 5 (2014), 1187–1199.

Collaborating with an Autonomous Agent to Generate Affective Music

5:21

- Fabio Morreale, Aliaksei Miniukovich, and Antonella De Angeli. 2014. Twiterradio: Translating tweets into music. In *Extended Abstracts on Human Factors in Computing Systems*. ACM. 738
- Fabio Morreale and Antonella De Angeli. 2015. Evaluating visitor experiences with interactive art. In *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter*. ACM, 50–57. 739
- Gerhard Nierhaus. 2009. *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer Science & Business Media. 740
- António Pedro Oliveira and Amílcar Cardoso. 2010. A musical system for emotional expression. *Knowledge-Based Systems* 23, 8 (2010), 901–913. 741
- Francois Paque. 2003. The continuator: Musical interaction with style. *Journal of New Music Research* 32, 3 (2003), 333–341. 742
- Philippe Pasquier, Arne Eigenfelt, and Oliver Bown. 2012. Preface. Musical Metacreation. In *Papers from the 2012 AIIDE Workshop AAAI Technical Report WS-12-16*. 743
- W. Piston. 1941. *Harmony*. W. W. Norton. 744
- Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 745
- Melvin G. Rigg. 1964. The mood effects of music: A comparison of data from four investigators. *Journal of Psychology* 58, 2 (1964), 427–438. 746
- Curtis Roads and John Strawn. 1985. *Foundations of Computer Music*. Vol. 28. MIT Press, Cambridge, MA. 747
- James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161. 748
- Emery Schubert. 1999. Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology* 51, 3 (1999), 154–165. 749
- Elliott Schwartz and Daniel Godfrey. 1993. *Music Since 1945: Issues, Materials, and Literature*. Schirmer Books, New York. 750
- Ian Simon, Dan Morris, and Sumit Basu. 2008. MySong: Automatic accompaniment generation for vocal melodies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 751
- Mark J. Steedman. 1984. A generative grammar for jazz chord sequences. *Music Perception* (1984), 52–77. 752
- Vygandas Šimbelis et al. 2014. Metaphone: Machine aesthetics meets interaction design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 753
- Peter M. Todd and Gregory M. Werner. 1999. Frankensteinian methods for evolutionary music. *Musical Networks: Parallel Distributed Perception and Performance* (1999), 313. 754
- Isaac Wallis et al. 2011. A rule-based generative music system controlled by desired valence and arousal. In *Proceedings of 8th International Sound and Music Computing Conference*. 755
- Gregory D. Webster and Catherine G. Weir. 2005. Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion* 29, 1 (2005), 19–39. 756
- Geraint Wiggins et al. 1998. Evolutionary methods for musical composition. evolutionary methods for musical composition. *International Journal of Computing Anticipatory Systems*. 757
- Ellen Winner. 1982. *Invented worlds: The Psychology of the Arts*. Harvard University Press. 758
- Hans T. Zeiner-Henriksen. 2015. Emotional experiences of ascending melodic lines. In *Proceedings of the 11th International Symposium on CMMR*. 759
- Marcel Zentner and Tuomas Eerola. 2010. Self-report measures and models. In *Handbook of Music and Emotion*. 187–221. 760

Received XXX; revised XXX; accepted XXX

Q23

QUERIES

- Q1:** AU: Please check author name and affiliation, which do not match with pdf.
Q2: AU: Please provide CCS 2012 Concepts per author guidelines and provide XML codes as well.
Q3: AU: Please provide complete mailing and email addresses for all authors.
Q4: AU: Please check “8/7” here—was “seven to eight” meant?
Q5: AU: Please check “couples of visitors”—was “a couple of visitors” meant?
Q6: AU: Please provide page numbers.
Q7: AU: Please provide page numbers.
Q8: AU: Please provide volume number.
Q9: AU: Please provide volume number.
Q10: AU: Please provide volume number.
Q11: AU: Please provide journal title and volume number.
Q12: AU: Please provide all authors’ names.
Q13: AU: Please provide publisher name.
Q14: AU: Please provide all authors’ names.
Q15: AU: Please provide all authors’ names.
Q16: AU: Please provide volume number.
Q17: AU: Please provide all authors’ names.
Q18: AU: Please provide volume number.
Q19: AU: Please provide all authors’ names.
Q20: AU: Please provide all authors’ names.
Q21: AU: Please provide volume and page numbers.
Q22: AU: Please spell out “CMMR”.
Q23: AU: Please provide article history dates.