

# Investigating Perceived Emotional Correlates of Rhythmic Density in Algorithmic Music Composition

---

DUNCAN WILLIAMS, ALEXIS KIRKE AND EDUARDO MIRANDA, Plymouth University

IAN DALY, JAMES HALLOWELL, JAMES WEAVER, ASAD MALIK, ETIENNE ROESCH, FAUSTINA HWANG, AND SLAWOMIR NASUTO, University of Reading

---

Affective algorithmic composition is a growing field that combines perceptually motivated affective computing strategies for novel music generation. This paper presents work towards the development of one such application. The long-term goal is to develop a responsive and adaptive system for inducing affect that is both controlled and validated by biophysical measures. Literature documenting perceptual responses to music identifies a variety of musical features and possible affective correlations, but perceptual evaluations of these musical features for the purposes of inclusion in a music generation system are not readily available. A discrete feature, rhythmic density, was selected on the basis that it was shown to be well-correlated with affective responses in existing literature. A prototype system was then designed to produce controlled degrees of variation in rhythmic density via a transformative algorithm. A two-stage perceptual evaluation of a stimulus set created by this prototype was then undertaken. First, listener responses from a pairwise scaling experiment were analysed via Multidimensional scaling (MDS). The statistical best-fit solution was rotated such that stimuli with the largest range of variation were placed across the horizontal plane in 2 dimensions. In this orientation, stimuli with deliberate variation in rhythmic density appeared further from the source material used to generate them than from stimuli generated by random permutation. Second, the same stimulus set was then evaluated according to the order suggested in the rotated 2-dimensional solution, in a verbal elicitation experiment. A verbal protocol analysis (VPA) found that the listener perception of the stimulus set varied in at least two commonly understood emotional descriptors, which might be considered affective correlates of rhythmic density. Thus, these results suggest that some parameterised control of perceived emotional content in an affective algorithmic composition system is possible, and provide a methodology for evaluating and including further possible musical features in such a system. Some suggestions regarding the test procedure and analysis techniques are also documented here.

---

## 1. INTRODUCTION

In the field of computer music research, perceptual models have recently been applied to algorithmic composition routines to develop new systems for the creation of affectively-charged, or affectively-driven music scores. This emerging field has been referred to as affective algorithmic composition (AAC) (Kirke and Miranda, 2011; Mattek, 2011; Williams et al., 2013). Whilst many such systems exist, research documenting the validation of affective mappings to isolated musical features by experiment is sparse. If AAC systems in the future intend to incorporate a full range of perceptual input, then a method for determining, and quantifying, the underlying musical features which might be used as perceptual correlates for this range of input is needed. This paper presents work towards this goal by implementing an isolated musical feature in a prototype AAC system and subjecting a stimulus set created by the prototype system to a perceptual evaluation. A literature review of emotional responses to musical features was carried out in order to determine possible perceptual correlates that might be included in an AAC system (Williams et al., 2013). The review determined that the psychological approaches to musical stimuli broadly relay three kinds of emotional responses to music as follows:

*Emotions: Short episodes; usually evoked by an identifiable stimulus event that may then further influence or direct the perception.*

*Affect/Subjective feeling: An experience of emotions or feelings (evoked in this case by music in the listener).*

*Moods: Longer-lived than emotions or affects, moods are a more diffuse construct and are usually latent and indiscriminate as to their eliciting events.*

The distinction between perceived and induced/experienced emotions has been well documented (see for example (Västfjäll, 2001; Vuoskoski and Eerola, 2011) (Gabrielsson, 2001)), though the precise terminology used to differentiate the two varies enormously. Perhaps unsurprisingly, results tying musical parameters to induced/experienced emotions do not provide a clear description of the mechanisms at play

(Juslin and Laukka, 2004; Scherer, 2004), and the terminology used can be inconsistent. Interested readers can find more exhaustive reviews on the link between music and emotion in (Scherer, 2004) and (Lamont and Eerola, 2011) .

Various techniques for reporting either perceived or experienced emotion in literature include self-reporting, and measurement of physical change, though neither are without their problems when applied to the parameterisation of emotion-laden music for affective induction. Firstly, bodily symptoms alone are not sufficient to evoke and consequently allow for the report of emotions (Schachter and Singer, 1962). Second, self-reporting techniques present challenges for the researcher who needs to measure affective phenomenon without disturbing or influencing the report in any way. Some research has confirmed that the same piece of music can elicit different responses, at different times in the same listener (Juslin and Sloboda, 2010). These challenges emphasise the variability of emotional phenomena, both in terms of signals that may be used as input to automatic systems to seed composition, and in terms of the range of creative output that may be available to an algorithmic composition system. These levels of complexity also suggest that a fully-fledged system should ultimately utilize many biophysical signals, in combination with reliable and accurate first-person reports, in order to accurately derive perceived emotional responses to musical stimuli from the system.

### 1.1 Affectively-driven algorithmic composition (AAC)

Algorithmic composition (either computer assisted or otherwise) is developing into a well-understood and documented field (Collins, 2009, 2009; Miranda, 2001; Nierhaus, 2009; Papadopoulos and Wiggins, 1999). Rowe (Rowe, 1992) describes three methodological approaches to algorithmic composition: generative, sequenced, or transformative. Discrete musical feature-sets, or rules for specific musical features, can be used as the input for algorithmic composition systems. One way of targeting affective responses by means of algorithmic composition would be to adapt affective measurement to the selection of rules for specific musical features in such a system.

Figure 1 provides an overview of the inputs and outputs an algorithmic composition system of this kind might use in order to produce an affective output. The system paradigm is that emotional correlates, determined by literature review or affective experimentation, can be used to inform the generative or transformative rules in order to target specific affective responses in the output. Previous research towards affective performance algorithms confirms that musical feature selection is unlikely to be trivial (Friberg et al., 2011). Regardless of feature choices, the feature-set must be implementable in some form by computer, and should likely have a known or expected emotional correlation in order to create an affective output. MIDI is often used as a musical data representation that can be both generated and/or transformed by computer, and many systems make use of MIDI as the musical data source.

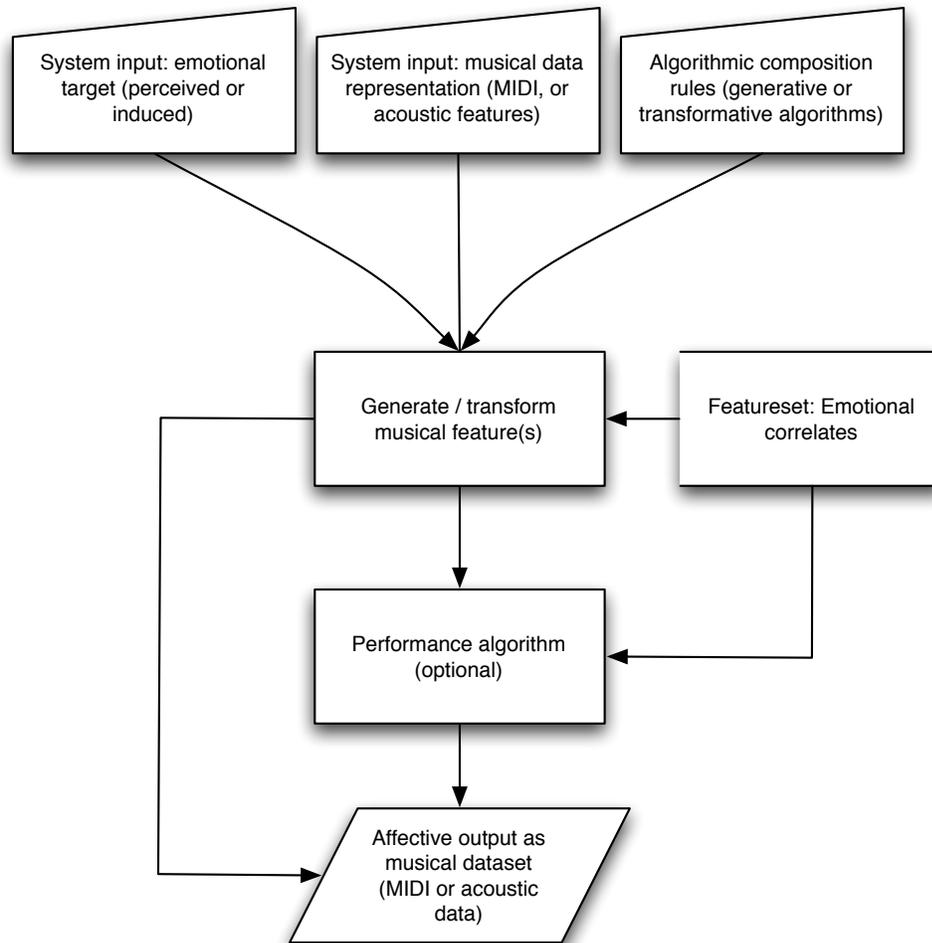


Fig. 1. Overview of basic affective algorithmic composition system using generation and manipulation of affectively correlated musical features as an input/control signal, including optional performance system. A minimum of three inputs are required: algorithmic compositional rules (generative, or transformative), a musical (or in some cases acoustic) dataset, and an emotional target.

1.1.1. *Why use discrete feature variation in a generative composition system?* Implementing and testing a composition system has various uses. First, it allows for confirmatory studies of existing emotional correlations to discrete musical features, providing an opportunity for individual feature evaluation in the context of generative or transformative automated composition. Without a rigorous listener evaluation it is difficult to be sure of the correct use of discrete musical feature control in such a system. Second, a composition system for the generation of novel material avoids some of the complications in affective evaluation that musical familiarity might give rise to (Marin and Bhattacharya, 2010). For example, if a listener is well used to a particular piece of music they may develop fatigue or frustration much sooner than if they are repeatedly exposed to 'new' music (Ladinig and Schellenberg, 2012). Furthermore, using existing music in such an experiment also runs the risk of exposing listeners to music they have already made emotional connections with as a contributory element of an episodic memory (imagine a participant being exposed to a piece of music with very sad, personal connotations, for example). These kinds of biases can potentially be circumnavigated by using original, algorithmically generated music. Finally, there is the possibility of creating a system for affective *induction*, if the system is adaptable to an individual listener's affective state. The distinction here is between cognitive understand and individual processing of affective state – for example, sad music may be deemed to be enjoyable by a listener in the appropriate state (Vuoskoski and Eerola, 2012; Vuoskoski et al., 2012). Therefore, this is a non-trivial adaptation given that the system would have to respond in near real-time to changes in affective state. It is likely that this would necessitate an affective matrix which accounts for the *relative* change induced by a whole range of particular music features (for example, going from a minor mode to a major mode may create a different

change if the listener already felt sad and had been enjoying the piece in the minor mode, they might be less satisfied with the change to the major mode).

*1.1.2 Specification challenges.* Whilst some musical features have a well-defined range of acoustic cues (pitch with fundamental frequency, tremolo with specific variation in amplitude envelope, etc.), others have more complicated acoustic (and/or musical) correlations. Therefore an awareness of the listeners' methods for perceiving these features, and any hierarchical interaction between such features becomes important when selecting an isolated musical feature for evaluation. Meter, for example (correlated with some emotions by (Kratus, 1993)), has been shown to be affected by both melodic and temporal cues (Hannon et al., 2004), as a combination of duration, pitch accent, and repetition (which might themselves then be considered 'low-level' features, with meter a 'higher-level', composite feature). Whereas pitch is more simply correlated with fundamental frequency. Many timbral features are also not clearly, or universally, correlated (Aucouturier et al., 2005; Bolger, 2004; Schubert and Wolfe, 2006), particularly in musical stimuli, presenting similar challenges to the selection and implementation of a single timbral feature for experimentation.

## 2. PROTOTYPE SYSTEM DESIGN

The methods which might be implemented in a system for manipulation of musical features by means of algorithmic composition vary depending on the feature in question. Therefore the issue of feature selection will be addressed before a discussion of the design and implementation of the system for manipulation of the selected feature.

### 2.1 Feature selection

In order to be useable in an AAC system, a given musical feature requires a method for *identification* and subsequent manipulation., e.g., *by using* structural or acoustical correlates. Features with affective correlations used in existing AAC systems include modality, rhythm, and melody, with 29, 29, and 28 instances respectively in a survey of recent work (Williams et al., 2013). These features can be considered to be groups of features that include an implicit hierarchy of sub-features. For example, pitch contour and melodic contour make a significant contribution to the instances of pitch features as a whole (Lerdahl and Jackendoff, 1983).

Of the two most popular feature groups, modality and rhythm, modality included 9 direct references and 20 references to sub-features (register, key, tonality etc). Rhythm included 11 direct references and 18 references to sub-features (meter, duration, time-signature etc). Therefore, rhythm appeared to be the most universally agreed upon feature-set included in existing AAC systems. However, for the purposes of the design of this prototype system, rhythm as a whole would be a difficult choice to implement as an isolated feature for perceptual evaluation due to the complex interactions between the sub-features and their respective contributory acoustic cues (Hannon et al., 2004). Therefore, the most common sub-feature of rhythm – rhythmic density – was selected for inclusion in the prototype, in order to minimize unwanted interaction from other musical features. Perceptual orthogonality with other features in the hierarchy cannot be assumed without experiment, not least as changes in rhythmic density afford various interesting musical and perceptual correlations. For example, a decrease in density can cause a change in perceived modality *as well as* in perceived tempo (even if the meter and pulse remain the same) – both of these changes can have a subsequent impact on the affective content of the music (Gagnon and Peretz, 2003).

### 2.2 System design

Rowe (Rowe, 1992) describes three methodological approaches to algorithmic composition: generative, sequenced, or transformative. Transformative approaches make use of existing musical material as an input – one or more transformations are applied to this input in order to create novel, yet related, material. A simple pitch inversion can be considered an example of a transformative process. One of the main differences between a transformative system and the generative or sequenced systems is that in a transformative system, the input signal need only contain musical data, rather than a control function of some sort. This lends transformative systems to techniques for 'aping' existing styles by process of deconstruction, analysis, and recombination. Transformative systems have an advantage over generative systems in that they contain a *de facto* rule-set, established by analysis of seed material. Thus, with a transformative system, there is no necessity to specify a large body of additional structural rules.

Given the musically successful implementation of other transformative systems (see for example, the *Experiments in Musical Intelligence* work of Cope (Cope and Mayer, 1996; Cope, 1992, 1989)), the prototype system presented here was designed to use a transformative algorithm in order to manipulate an isolated musical feature (rhythmic density, as mentioned above, which is a temporal aspect of music derived from pulses and meter, contributing to perceived tempo, and to a lesser extent, perceived modality). If the transformative prototype is found to achieve perceptual variation with such a limited musical rule-set, then in the future it might be further adaptable to generative operation via the addition of extra structural rules. This will facilitate work towards a larger system for AAC based on selective manipulation of a broad range of musical features and the targeted, underlying emotional correlations.

An offline, transformative system was prototyped in OpenMusic (Bresson et al., 2005) and common Lisp. The signal flow of the prototype system is illustrated in Figure 2. The prototype system was designed to utilize monophonic data (i.e., single handed piano, stringed instruments or similar timbres).

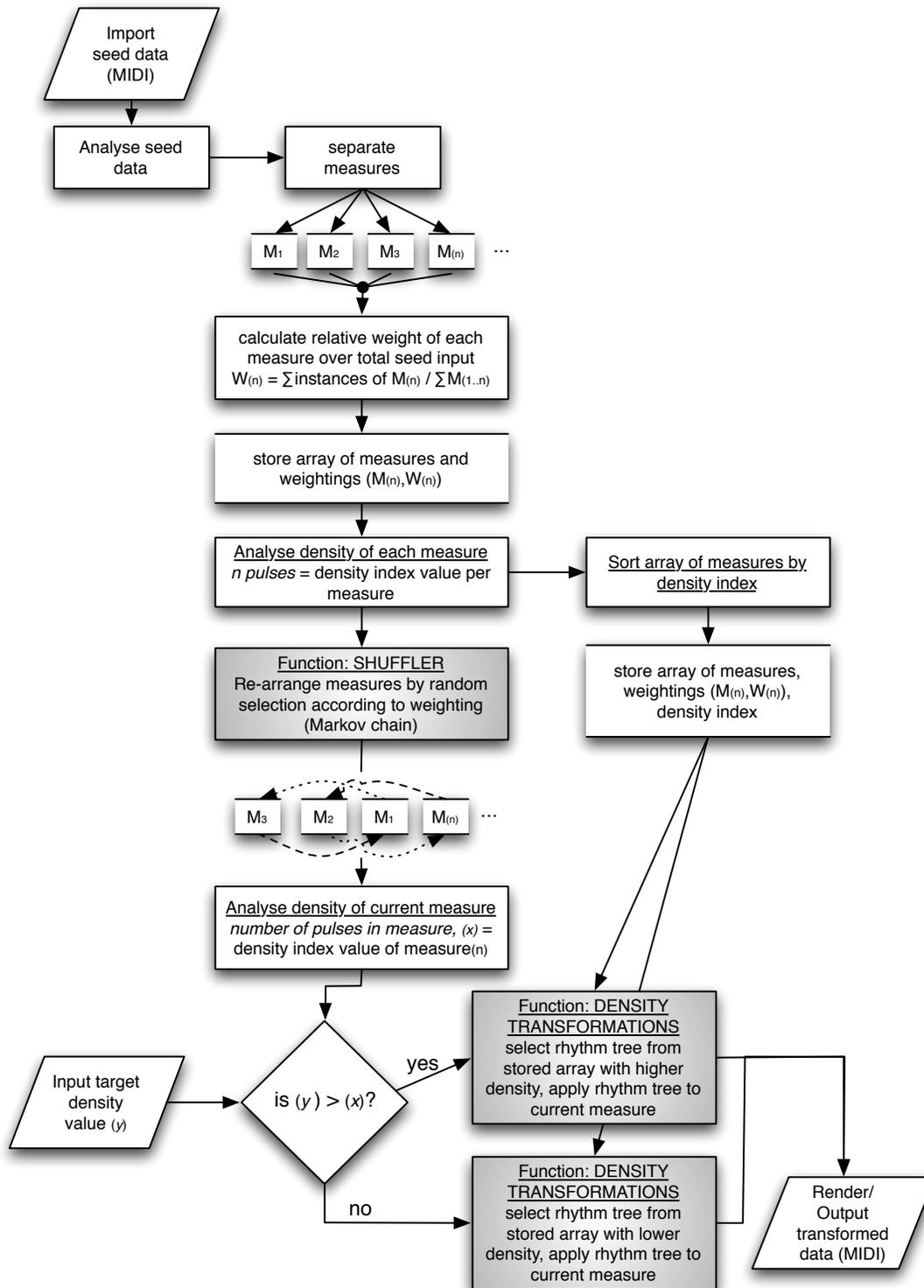


Fig. 2. Signal flow of the prototype system. It is possible to generate various permutations via this system, including the generation of a permuted set of measures using existing rhythm trees, a new set of measures with increased density (number of pulses extracted from other measures), and a new set of measures with decreased density.

The prototype uses three phases; learning (analysis), algorithmic transformation, and generation (rendering). At the learning phase, the system analyses a selected seed input by separating the musical

structure into measures and deriving a second-order transition matrix of pitch and rhythm tree information. This results in a stored hierarchical list representing rhythmic structures with probability values for the transitions between these structures. An array of the resulting values is then stored. A statistical analysis of rhythmic density is carried out on the stored array by searching for the number of pulses in each measure according to note onset and duration values. At the transformation phase, the derived density value is used as an index from which to create new permutations via a Markov chain of pitch and rhythm tree information using the transition matrix. Provided that there is enough variation in the original input material, new permutations can be created solely from measures with low-density index values, high-density index values, or a combination of the two. The transformed permutations are then rendered in the generation phase to allow the output to be saved as a MIDI format file for immediate playback, or subsequent editing. This prototype could be expanded by increasing the order of the Markov chain to incorporate more complex transitions, other musical features and higher-level musical structures in the future.

### 3. PERCEPTUAL EVALUATION

A two-stage experiment was devised in order to evaluate a stimulus set created using the prototype system. This experiment aimed to:

- (1) Select a perceptually meaningful and affectively correlated musical feature with which to test the existing documented correlations between the chosen feature and its perceived affective content.
- (2) Develop and evaluate a methodology for quantifying the perceived affective content created by the prototype such that it might be expanded to a larger, multi-feature AAC system in future.

First, a pairwise dissimilarity experiment was designed to test the best-fit number of perceived dimensions and construct a perceptual space for the stimulus set within that number of dimensions. The construction of a perceptual space using Multidimensional Scaling Analysis (MDS) from a set of listener evaluations has previously been shown to be a useful way to construct statistically meaningful dimensional models of listener perceptions of musical stimuli (E. Bigand et al., 2005; Emmanuel Bigand et al., 2005; Vieillard et al., 2008; Wu and Jeng, 2008). Confidence in such models can be evaluated by statistical measures in order to determine the best-fit dimensionality for the model, and to create a plot of the stimuli which shows the respective and relative similarities amongst the stimuli in the model. With MDS analysis, dimensional labels cannot be established in this process. The second stage of the perceptual evaluation therefore sought to determine perceptual labels for the dimensions revealed by the analysis of the first experimental stage. Stimuli were presented to the listeners in the second stage in the order that they were arranged in the best-fit perceptual space from the first stage, with the aim being to provide meaningful labels for any perceived movement in each of the resulting dimensions.

#### 3.1 Stimulus set generation

Stimuli for both stages of the experiment were created using the prototype system and 4 seed inputs from a previous study evaluating affective responses and neurophysical correlations in electroencephalogram (EEG) to western classical music (Schmidt and Trainor, 2001). These seed inputs were selected with the partial intention of adapting a brain computer musical interface (BCMI) system to the control of AAC via EEG in future. Thus, seed material that had already been perceptually evaluated by means of EEG was selected as a useful starting point. The sources and corresponding affective evaluations from Schmidt and Trainor (Schmidt and Trainor, 2001) were as follows:

*Peter and the Wolf (Prokofiev) - 2.05 pleasantness (valence), 6.18 intensity (arousal)*

*Brandenburg Concerto No. 5 (J.S. Bach) - 8.27 pleasantness (valence), 3.59 intensity (arousal)*

*Four Seasons: Spring (Vivaldi) - 7.91 pleasantness (valence), 2.45 intensity (arousal)*

*Adagio for Strings (Barber) - 2.91 pleasantness (valence), 1.91 intensity (arousal)*

Figure 3 shows an excerpt from the seed material before it had been separated into measures. Figure 4 shows a lower density excerpt generated by the prototype system from the same seed material. The seed material in this case was an excerpt from J.S. Bach's *Brandenburg Concerto No. 5*, as per Schmidt and Trainor, which consists mostly of 1/16<sup>th</sup> notes, with the exception of the material in the latter half of the figure. When the density transformation seeks to find material with lower density than the current measure, it uses the rhythmic tree suggested by this lower density material as a template from which to create new permutations of the material in the lower density output, and *vice versa* for high density transformation and generation. The score itself is not optimised by this routine and could be further edited by hand for ease of

sight reading, but it provides suitable material for immediate machine playback and thus for the generation of audio stimuli for subsequent perceptual evaluation by listener testing.



Fig. 3. Excerpt of seed material which has been condensed to a monophonic piano arrangement, taken from *Brandenburg Concerto No. 5*, J.S. Bach.



Fig. 4. 'Lower density' excerpt created by Markov permutation of measures from the seed material, with the low density index used as the basis for the selection of rhythm trees. Note that the algorithm has made use of triplets to emulate the pattern from the latter half of the seed material.

With MDS analysis, a minimum of 4 stimuli per dimension that can be revealed in the final analysis is required. Therefore, in order to allow for up to 4 dimensions of variation in the stimuli generated by the prototype system, 16 stimuli were prepared from the 4 seed inputs. The complete stimulus set was then as follows:

- 1-4: original material, edited for duration
- 5-8: lower density rhythmic transformations applied to seed material
- 9-12: higher density rhythmic transformations applied to seed material
- 13-16: permutations of original (Markov shuffling) with no rhythmic transformations

All stimulus material was limited to the same duration and condensed to monophonic playback via a piano timbre (Type 0 MIDI file). Changes in rhythmic density can have a knock-on effect on perceived mode, by, for example, creating a lower density generation which does not include the necessary pitches which correspond to a minor mode, thereby creating an ambiguous modality which might be major or minor. The same effect can also be noticed in perceived tempo, whereby a rhythmically sparse passage might appear to have a slower tempo than a rhythmically dense passage at the same actual BPM (Raphael, 2001; Whiteley et al., 2007). Table 1 shows the complete stimulus set including estimated tempo and mode values derived by an automated listener, ARTHUR (Kirke et al., 2013). Note that perceived mode and tempo can vary as functions of the change in rhythmic density, as in 12 and 16 where perceived mode changes from minor to major in the higher density transformation.

Table 1 Stimulus set used in pairwise dissimilarity comparison experiment, number, label, and type of processing used in stimulus preparation/generation..

Stimulus number/label	Content	Automatically estimated perceived tempo	Automatically estimated perceived mode (0= minor, 1 = major)
1. AdagioE	Adagio for strings edit	1.1318	0
2. BrandE	Brandenburg concerto edit	5.375	1
3. SpringE	Four seasons (spring) edit	2.61	0
4. WolfE	Peter and the Wolf edit	3.2323	1
5. AdagioLD	Adagio for strings lower density	0.5426	0

	transformation		
6. BrandLD	Brandenburg concerto lower density transformation	5.8636	1
7. SpringLD	Four seasons (spring) lower density transformation	0.9877	0
8. WolfLD	Peter and the Wolf lower density transformation	1.2530	1
9. AdagioHD	Adagio for strings higher density transformation	0.9468	0
10. BrandHD	Brandenburg concerto higher density transformation	7.2967	1
11. SpringHD	Four seasons (spring) higher density transformation	2.3457	0
12. WolfHD	Peter and the Wolf higher density transformation	1.6703	1
13. AdagioP	Adagio for strings permutation (no rhythmic transformations)	0.6277	0
14. BrandP	Brandenburg concerto permutation (no rhythmic transformations)	7.2273	1
15. SpringP	Four seasons (spring) permutation (no rhythmic transformations)	1.2593	0
16. WolfP	Peter and the Wolf permutation (no rhythmic transformations)	1.6386	0

### 3.2 Pairwise scaling stage

Twenty-two listeners participated in the first stage. Each participant had some experience of critical listening (all participants were in the third and final year of undergraduate study in music technology). Ethical approval for the experiment was granted by the Humanities and Performing Arts research committee of Plymouth University. All participants were aged between 22-35 and received no financial incentive to take part in the experiment. Two of the participants were female.

The experiment was conducted in a laboratory context simultaneously on 22 standalone desktop machines, each running a discrete version of the test interface. Circumaural headphones were used. Participants were allowed to adjust volume levels according to their own preference during a familiarization exercise. A reasonable acoustic isolation was achieved with screening between each workstation. The familiarization exercise also allowed listeners to hear the full range of stimuli in a non-linear fashion before undertaking the main experiment.

The scaling itself asked listeners to evaluate each of the stimuli against one another in 136 randomly ordered pairs, split over two tests of approximately 35 minutes in duration. In each comparison, listeners were asked to rate the similarity between a pair using a 100 point continuous scale with end-points labeled 'not at all similar' and 'the same', as shown in Figure 5 (the middle point of the scale was labeled 'fairly similar').



Fig. 5. Screenshot of a single evaluation from the *Max/MSP* listener interface used in the pairwise scaling experiment. Listeners are invited to evaluate the similarity between A and B using the slider on a pair-by-pair basis, eventually comparing every stimulus from the set.

**3.2.1 Pairwise scaling results.** Listener responses to the pairwise scaling were collated to produce a dissimilarity matrix which was then subjected to an Individual Differences Scaling (INDSCAL) MDS analysis (Kruskal, 1964) in order to establish the number of dimensions that best represented the variation listeners had perceived across the stimulus set. The statistical 'measures-of-fit' determined by the analysis (dimensionality, RSQ or square of the correlation coefficient, and Kruskal stress) are shown in Table 2.

Table 2 Statistical 'measures-of-fit' determined by MDS INDSCAL analysis of listener responses. Measures in bold indicate a quality criterion has been met. The maximum possible RSQ improvement at 4-Dimensions is given by 1-(4-D RSQ).

Dimensionality	RSQ	RSQ improvement in next increase in dimensionality	Stress (Kruskal stress formula 1)
1-D	<b>0.99914</b>	<b>0.00067</b>	0.574
<b>2-D</b>	<b>0.99981</b>	<b>0.00001</b>	<b>0.200</b>
3-D	<b>0.99982</b>	<b>0.00014</b>	<b>0.109</b>
4-D	<b>0.99996</b>	n/a	<b>0.072</b>

In any MDS analysis, an increase in the number of dimensions utilized by the solution will decrease the amount of stress, hence determining the optimum solution is not simply a matter of looking for the lowest stress. The statistical measures in Table 1 were examined to determine the 'correct' dimensionality as the number of dimensions which best represented the perceived variation in the stimulus set. The criteria which could be used as indicators of statistical quality include RSQ greater than 0.95 (Astill, 1994), stress greater than 0.20 and optimally as low as 0.05 (Kruskal, 1964), and a negligible improvement in RSQ at a given increase in dimensionality. Table 1 shows that RSQ was greater than 0.95 in all dimensionalities, suggesting that each was a confident solution. The RSQ improvement at each additional dimension was also low, though the lowest improvement is found between the 2 and 3-Dimensional solutions. Stress was highest in the 1-Dimensional solution, but below the threshold of <0.20 in all other solutions. An examination of the scree plot showing stress against dimensionality found a significant knee (which can also be interpreted as an indicator of 'correct' dimensionality), at 2-Dimensions, as shown in Figure 6. Together, these results strongly suggested a 2-D solution. The spread in a Shepard diagram at 2-Dimensions, as shown in Figure 7, was also examined, with a low spread in the data confirming a statistically good fit at this dimensionality.

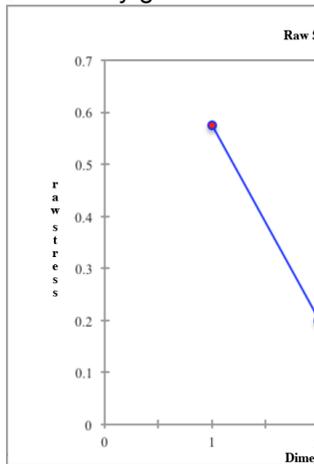
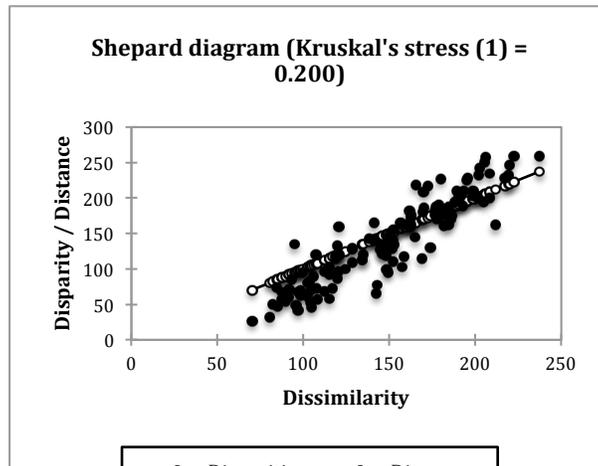


Fig. 6. Scree plot showing a significant knee in the results at 2 dimensions, with stress at



significant knee in the results 0.200.

Fig. 7. Shepard diagram showing a low spread in the results between the similarities and distances in the 2-D solution, with stress at 0.200.

With a statistically confident solution, the perceptual space could then be plotted in 2-Dimensions.

3.2.2. *Perceptual space in 2-D.* The arrangement of stimuli in Figure 8 can be considered a perceptual space, via which listener responses to the stimulus set can be explored. In a similar fashion to the use of timbre space as a control structure for sound synthesis (Wessel, 1979), this perceptual space could, if expanded, be used as a control structure for the generation of affectively-charged musical stimuli, for example by means of algorithmic composition techniques in future. The 2-Dimensional spacing suggests that the permuted stimuli are perceptually closer to the respective seed material than the corresponding density transformations, which indicates that a permutation in overall musical structure may have less perceptual significance to listeners than an isolated variation in rhythmic density. In other words, even when the output is modified significantly by the process of random permutation, the resulting music retains more perceptual similarity to the seed material than the output generated by selectively and deliberately manipulating rhythmic density in isolation.

Two anomalies are present in the 2-Dimensional perceptual space. The 'Adagio' group (from *Adagio for Strings* by Barber) appears to show the placing of the seed stimulus and the high density transformation in positions which do not follow the general trend. This might be explained by the significantly lower density found in the Adagio seed material, which is a slow, sparse piece of music in comparison to the other input seed sources. The 'Brand' group (from *Brandenburg Concerto No. 5* by J.S. Bach) also exhibits some unusual placing in the perceptual space. In this group, although the permuted stimulus remains the closest to the original seed, the density transformations are positioned atypically. The seed material for this group is considered to be the 'most dense' by the prototype system, with the largest number of onsets and shortest durations. This might explain why the 'Brand' group is presented approximately opposite the 'Adagio' group, and also why the listeners perceived the variation in the atypical manner plotted in Figure 8. However, if the angle of the configuration is rotated whilst still maintaining the direction of perceived density in other seed groups from left to right (i.e., with low to high instead of high to low in dimension 2), the stimuli in question then appear to be ordered BrandLD, BrandP, BrandE, and BrandHD, as would be expected according to the general trends observed above.

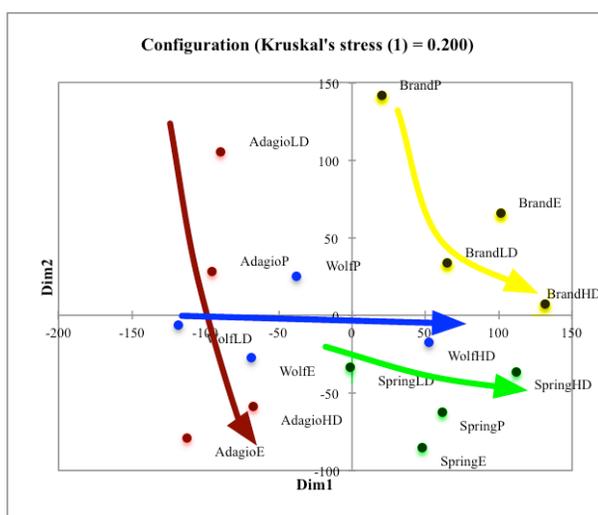


Fig. 8. Perceptual space in 2-Dimensions after MDS INDSCAL (individual differences scaling) analysis. Movement can be seen across the direction of low to high density stimuli. Colored annotations show the grouping of stimuli based on the seed material used in the learning phase. Stimuli appended -E are the original edited seed excerpts. Stimuli appended -P are the permutations with no intended change in rhythmic density. Stimuli appended -LD are the low density transformations, and stimuli appended -HD are the high density transformations.

The perceptual space shows that the transformed stimuli are loosely grouped nearby to their seed material, with a general trend that low density transformations are found in the upper left of the corresponding seed group, and high density transformations in the lower right of the seed group. Overall there is a tendency for an increase in density to be plotted across the perceptual space from the upper left of the space to the lower right. This spacing bears a similarity to some existing work using the circumplex model of affect (Russell, 1980), a 2-Dimensional emotional space whose dimensions correspond to affective arousal and valence, which has been adapted to the psychological evaluation of music and to affectively-charged algorithmic composition in some systems (Kirke and Miranda, 2011; Wallis et al., 2011). Whilst such observations can only be casually drawn, Barber's *adagio* seems to be an anecdotally 'sadder', more somber piece, whilst the Brandenburg concerto is faster, more lively, higher energy and subjectively 'happier', corroborated by a position at the opposite end of the 2-D space to that of the *adagio* derived stimuli. This strongly suggests that isolated musical feature manipulation is compatible with this method of parameterizing affect for algorithmic control, and that in the future, a larger system, incorporating several isolated features as part of an AAC system should be possible.

Subsequently, the second stage of perceptual evaluation, a subsidiary verbal elicitation experiment, was undertaken.

### 3.3 Verbal elicitation stage

MDS analysis cannot reveal the names of the dimensions given by the pairwise dissimilarity experiment, and a subsidiary verbal elicitation experiment was undertaken to establish whether listeners perceived the changes in density to correlate to any specific changes in musical affect.

Listeners were invited to review pairs of stimuli and describe any perceived changes from left to right in the pair, using as many adjectives as they felt necessary in order to fully describe the change. Emotional responses were not specifically demanded by the interface. In order to shorten the test duration, the permuted stimuli were discarded as they were plotted next to the source stimuli in each group in the 2-D perceptual space. Therefore, each source sound was compared separately with its own low density and high density generations, making 8 total comparisons. 8 participants, all experienced listeners, were presented with a *Max/MSP* interface to undertake the evaluation, using the same circumaural headphones and in the same venue as in the pairwise dissimilarity experiment.

**3.3.1 Verbal elicitation results.** Listener responses were grouped together based on their meaning (synonyms, antonyms and any other commonalities), by an independent academic with prior experience of conducting similar groupings, but with no knowledge of the experiment's aims. The number of listener responses in each group was summed and an overall prominence, indicating the perceptual importance of each group, was calculated by dividing the number in each group by the total number of responses. The groupings, number of occurrences, and overall prominence for each set of comparisons are shown in Table 3 and 4.

Table 3 Adjective groupings, occurrences, and overall prominence of verbal descriptors used when describing change from source to low density

Similar adjectives used to describe change across group	Number of occurrences (total: 32)	Overall prominence (occurrences / total)
Sombre, solemn, serious, calm, sad	12	0.375
Brooding, gloomy, dark	8	0.25
Longing, lonely	6	0.1875
Soothing, dreamy	4	0.125
Slow	4	0.125
Dull	2	0.0625
Suspicious	1	0.03125
Legato	1	0.03125

Table 4 Adjective groupings, occurrences, and overall prominence of verbal descriptors used when describing change from source to high density

Similar adjectives used to	Number of occurrences	Overall prominence
----------------------------	-----------------------	--------------------

describe change across group	(total: 38)	(occurrences / total)
Humorous, quirky, playful, cheerful, happy, perky	15	0.39473
Fast/Faster	10	0.26315
Uplifting, soaring, flying	6	0.15789
Majestic, proud, regal	5	0.13157
Stacatto, spiky	2	0.05263

The adjectives shown in Table 3 and Table 4 comprise a range of descriptors, including some direct musical features (slow, legato, staccato) and some well-known perceptual descriptors (happy, sad, solemn). Direct musical feature description was more common when describing the High Density stimuli, with 12 instances of Fast/Faster, and Stacatto/Spiky, whilst the Low Density stimuli only attracted five instances of such descriptors (slow and legato). The most prominent descriptors, based on overall prominence, are perceptual:

*Low density descriptors: Sombre, solemn, serious, calm, sad*

*High density descriptors: Humorous, quirky, playful, cheerful, happy, perky*

All of the prominent low density comparison descriptors, with the exception of calm, have been previously mapped directly on to dimensions 1 and 2 of Hevner's adjective cycle which describes elements of emotional expression in music (Hevner, 1936). All of the prominent high density descriptors, with the exception of quirky and perky, are already mapped on to dimensions 5 and 6 of the adjective cycle. This suggests that there is a strong emotional correlation in movement across the perceptual space in Figure 8 between increased density and an increase in arousal and valence, specifically giving listeners the impression of a move from sombre to humourous, or sad to happy and so on.

#### 4. DISCUSSION

Together, these experiments suggest that listener responses to algorithmically transformed music can be used to determine perceptual, and specifically, perceived emotional correlations to a single musical feature, (in this case, rhythmic density). If other musical features can be mapped to this perceptual space with strong affective correlations, an emotional control space for musical feature generation and manipulation could be developed in order to target affective responses through music. However, quantifying perceptual overlap amongst musical features remains a significant area for further work before such descriptors could be used as a 'musical control structure' in an AAC system.

The size and scale of these experiments, in particular the pairwise scaling stage, suggests that this methodology would not necessarily be the best way forward for evaluating perceptual overlap in musical feature manipulation – the orthogonal movement in the 2-D perceptual space used in this analysis is already difficult to interpret visually. Higher dimensionalities typically result in more complex visual plots, and adding new musical features is likely to increase the dimensionality of the perceptual space significantly. The [verbal protocol analysis \(VPA\)](#) suggests that qualitative experiments might be a more appropriate methodology for further work: the VPA revealed a good deal of listener corroboration and only a relatively small number of non-affective descriptors were found in the results. Nevertheless, these responses effectively represent 'wasted' data if attempting to determine solely emotional correlations. An affective interface for emotional descriptors (for example, a fixed or multiple choice profile) might result in less wasted data whilst evaluating more complex systems in future.

#### 5. CONCLUSIONS

A fully realized AAC system would have an advantage over traditional algorithmic composition in that it should be able to generate affectively-charged musical structures automatically and reactively, in response to a user's emotional state. In order to determine whether isolated musical features could be used in such a system, a prototype for generating new musical structures from seed material with varying levels of rhythmic density was developed and evaluated by means of a two-stage perceptual experiment.

The pairwise dissimilarity stage concluded that listener responses to stimuli generated by the prototype could be plotted in a 2-D solution with reasonable statistical confidence, based on Kruskal stress 1, RSQ, [and relative improvement in RSQ](#) across four dimensions, a significant knee in the scree plot at 2-D, and a low spread in a Shepard plot of the data at 2-D. Within the [2-dimensional](#) space, randomly

permuted stimuli were shown to be most perceptually similar to the seed material, with less perceived similarity to stimuli created with deliberate variation in rhythmic density. This is a somewhat surprising finding and has implications for the incorporation of a larger range of isolated musical features in an AAC system. The exact position of each stimulus in the 2-D plot suggested that there was some difficulty in successfully manipulating source material that was initially very musically sparse or dense, but otherwise the 2-D space suggested that stimuli were perceived as intended, in an order which suggested clearly perceptible movement from low to high rhythmic density. The 2-Dimensional perceptual space also showed a marked similarity to the 2-Dimensional model of affect which some algorithmic composition systems have adopted in order to automatically generate emotionally charged music. This further suggests that additional isolated feature manipulation could successfully contribute to a larger system for affect-driven algorithmic composition in the future. However, the complexity and relatively high stress of the 2-D space suggests that this particular method would not be best suited to future analysis involving larger numbers of perceptually overlapping musical features, as this would likely increase dimensionality and stress in the subsequent solutions, rendering a perceptual space which would be very difficult to interpret. Furthermore, the degree of control over the perceptual unidimensionality in the correlations noted above is, to some extent, dependent on the initial density of the seed material, which was itself limited to a small range from the western classical repertoire.

The subsequent verbal elicitation experiment, aiming to label the changes in the 2-D perceptual space found by the scaling stage, suggested that rhythmic density was well correlated with a number of affective descriptors, which, when grouped according to the number of instances and their perceptual prominence, suggest density is correlated to both arousal and valence, with low density stimuli being perceived as sad, solemn, and serious, whilst high density stimuli were perceived as happy, humorous, and playful.

Emotional descriptors that have been found to be correlated with changes in density by these experiments could now be used in an affective algorithmic composition system as a **structural** control. Within the caveats outlined above, these results suggest that more complex affective algorithmic composition systems, using a range of emotional correlations as musical control structures should be viable in future.

## 6. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of EPSRC grants EP/J003077/1 and EP/J002135/1.

## REFERENCES

- Astill, B., 1994. An investigation of social values in a senior secondary school milieu, in: In Proceedings of the Australian Association for Research in Education Conference. Presented at the (AARE) Conference, University of Adelaide, Australia.
- Aucouturier, J.-J., Pachet, F., Sandler, M., 2005. "The way it Sounds": timbre models for analysis and retrieval of music signals. *IEEE Trans. Multimed.* 7, 1028–1035. doi:10.1109/TMM.2005.858380
- Bigand, E., Filipic, S., Lalitte, P., 2005. The time course of emotional responses to music. *Ann. N. Y. Acad. Sci.* 1060, 429–437.
- Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., Dacquet, A., 2005. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cogn. Emot.* 19, 1113–1139. doi:10.1080/02699930500204250
- Bolger, D., 2004. Computational Models of Musical Timbre and the Analysis of its structure in Melody, in: PhD Thesis, University of Limerick.
- Bresson, J., Agon, C., Assayag, G., 2005. Openmusic 5: A cross-platform release of the computer-assisted composition environment, in: 10th Brazilian Symposium on Computer Music, Belo Horizonte, MG, Brésil.
- Collins, N., 2009. Musical Form and Algorithmic Composition. *Contemp. Music Rev.* 28, 103–114. doi:10.1080/07494460802664064
- Cope, D., 1989. Experiments in musical intelligence (EMI): non-linear linguistic-based composition. *J. New Music Res.* 18, 117–139.
- Cope, D., 1992. Computer modeling of musical intelligence in EMI. *Comput. Music J.* 69–83.
- Cope, D., Mayer, M.J., 1996. Experiments in musical intelligence. AR Editions Madison, WI.
- Friberg, A., Schoonderwaldt, E., Hedblad, A., 2011. Perceptual ratings of musical parameters. *Gemessene Interpret.-Comput. Aufführungsanalyse Im Kreuzverhör Diszip.* Mainz Schott.
- Gabrielsson, A., 2001. Emotion perceived and emotion felt: Same or different? *Music. Sci. Spec Issue*, 2001-2002, 123–147.
- Gagnon, L., Peretz, I., 2003. Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cogn. Emot.* 17, 25–40.
- Hannon, E.E., Snyder, J.S., Eerola, T., Krumhansl, C.L., 2004. The Role of Melodic and Temporal Cues in Perceiving Musical Meter. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 956–974. doi:10.1037/0096-1523.30.5.956
- Hevner, K., 1936. Experimental studies of the elements of expression in music. *Am. J. Psychol.* 48, 246–268.
- Juslin, P.N., Laukka, P., 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238.
- Juslin, P.N., Sloboda, J.A., 2010. *Handbook of music and emotion : theory, research, applications.* Oxford University Press, Oxford.
- Kirke, A., Miranda, E., 2011. Combining EEG Frontal Asymmetry Studies with Affective Algorithmic Composition and Expressive Performance Models.

- Kirke, A., Miranda, E., Nasuto, S.J., 2013. Artificial Affective Listening towards a Machine Learning Tool for Sound-Based Emotion Therapy and Control, in: Proceedings of the Sound and Music Computing Conference. Presented at the Sound and Music Computing, SMC, SMC Network, Stockholm, Sweden, pp. 259–265.
- Kratus, J., 1993. A developmental study of children's interpretation of emotion in music. *Psychol. Music* 21, 3–19.
- Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27.
- Ladinig, O., Schellenberg, E.G., 2012. Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychol. Aesthet. Creat. Arts* 6, 146.
- Lamont, A., Eerola, T., 2011. Music and emotion: Themes and development. *Music. Sci.* 15, 139–145. doi:10.1177/1029864911403366
- Lerdahl, F., Jackendoff, R., 1983. An overview of hierarchical structure in music. *Music Percept.* 229–252.
- Marin, M.M., Bhattacharya, J., 2010. Music induced emotions: Some current issues and cross-modal comparisons.
- Mattek, A., 2011. Emotional Communication in Computer Generated Music: Experimenting with Affective Algorithms, in: Proceedings of the 26th Annual Conference of the Society for Electro-Acoustic Music in the United States. Presented at the SEAMUS, University of Miami Frost School of Music, Miami, Florida.
- Miranda, E.R., 2001. *Composing music with computers*, 1st ed. ed, Music technology series. Focal Press, Oxford ; Boston.
- Nierhaus, G., 2009. *Algorithmic composition paradigms of automated music generation*. Springer, Wien; New York.
- Papadopoulos, G., Wiggins, G., 1999. AI methods for algorithmic composition: A survey, a critical view and future prospects, in: AISB Symposium on Musical Creativity. pp. 110–117.
- Raphael, C., 2001. *Automated Rhythm Transcription.*, in: ISMIR.
- Rowe, R., 1992. *Interactive music systems: machine listening and composing*. MIT press.
- Russell, J.A., 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161.
- Schachter, S., Singer, J., 1962. Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* 69, 379–399. doi:10.1037/h0046234
- Scherer, K.R., 2004. Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them? *J. New Music Res.* 33, 239–251. doi:10.1080/0929821042000317822
- Schmidt, L.A., Trainor, L.J., 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn. Emot.* 15, 487–500.
- Schubert, E., Wolfe, J., 2006. Does Timbral Brightness Scale with Frequency and Spectral Centroi... *Acta Acust. United Acust.* 92, 820–825.
- Västfjäll, D., 2001. Emotion induction through music: A review of the musical mood induction procedure. *Music. Sci. Spec Issue*, 2001-2002, 173–211.
- Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., Bouchard, B., 2008. Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cogn. Emot.* 22, 720–752.
- Vuoskoski, J.K., Eerola, T., 2011. Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Music. Sci.* 15, 159–173. doi:10.1177/1029864911403367
- Vuoskoski, J.K., Eerola, T., 2012. Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. *Psychol. Aesthet. Creat. Arts* 6, 204.
- Vuoskoski, J.K., Thompson, W.F., McIlwain, D., Eerola, T., 2012. Who enjoys listening to sad music and why? *Music Percept.* 29, 311–317.
- Wallis, I., Ingalls, T., Campana, E., Goodman, J., 2011. A Rule-Based Generative Music System Controlled by Desired Valence and Arousal. *SMC*.
- Wessel, D.L., 1979. Timbre Space as a Musical Control Structure. *Comput. Music J.* 3, 45–52.
- Whiteley, N., Cemgil, A.T., Godsill, S., 2007. Sequential inference of rhythmic structure in musical audio, in: *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on.* IEEE, pp. IV–1321–IV–1324.
- Williams, D., Kirke, A., Miranda, E.R., Roesch, E.B., Nasuto, S.J., 2013. Towards Affective Algorithmic Composition, in: Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th-15th June 2013. Geoff Luck & Olivier Brabant (Eds.). ISBN 978-951-39-5250-1. University of Jyväskylä, Department of Music.
- Wu, T.-L., Jeng, S.-K., 2008. Probabilistic estimation of a novel music emotion model, in: *Advances in Multimedia Modeling*. Springer, pp. 487–497.