Affective calibration of musical feature sets in an emotionally intelligent music composition system

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

Affective Calibration of Musical Featuresets in an Emotionally Intelligent Music Composition System

DUNCAN WILLIAMS, ALEXIS KIRKE AND EDUARDO MIRANDA, Plymouth University

IAN DALY, FAUSTINA HWANG, JAMES WEAVER, AND SLAWOMIR NASUTO, University of Reading

Affectively-driven algorithmic composition (AAC) is a rapidly growing field which exploits computer-aided composition in order to generate new music with particular emotional qualities or affective intentions. An AAC system was devised in order to generate a stimulus set covering 9 discrete sectors of a 2-dimensional emotion space by means of a 16 channel feed-forward artificial neural network. This system was used to generate a stimulus set of short pieces of music, which were rendered using a sampled piano timbre and evaluated by a group of experienced listeners who ascribed a two-dimensional valence-arousal co-ordinate to each stimulus. The underlying musical feature set, initially drawn from literature, was subsequently adjusted by amplifying or attenuating the quantity of each feature in order to maximize the spread of stimuli in the valence-arousal space before a second listener evaluation was conducted. This process was repeated a third time in order to maximize the spread of valence-arousal co-ordinates ascribed to the generated stimulus set in comparison to a spread taken from an existing pre-rated database of stimuli, demonstrating that this prototype AAC system is capable of creating short sequences of music with a slight improvement on the range of emotion found in a stimulus set comprised of real-world, traditionally composed musical excerpts.

1. INTRODUCTION

Affectively-driven algorithmic composition (AAC) is a rapidly growing field which exploits computer-aided composition in order to generate new music with particular emotional qualities or affective intents (Mattek, 2011; Duncan Williams et al., 2014), to create music which targets particular emotional responses in the listener. The intention of this intelligent type of system is to make music which adequately expresses a mood, where 'jukebox' selection (Eaton et al., 2014) of existing music by means of Music Information Retrieval (MIR) which typically uses acoustic analysis to extrapolate meaning from audio samples might not be suitable (e.g., in therapeutic applications). In order for such systems to be useful, they must be capable of creating a theoretically infinite number of pieces without time constraints. The need for this is dependent on possible end-use cases, for example therapeutic applications wherein hearing the same piece twice could cause unexpected or unpredictable emotional responses, or in commercial applications, e.g., gaming, wherein music would need to be generated in response to a continuously varying emotional trajectory dictated by gameplay cues (i.e., there is no saying when a particular piece of music may need to end or change in emotional content as this is dictated in real-time by the player). Repetition in video games soundtracking has previously been shown to be detrimental to immersion (Brown and Cairns, 2004; Grimshaw et al., 2008). This creates a number of challenges for a music generation system, particularly one which intends to target specific affective responses, as overall musical structure has also been shown to be a significant emotional communicator (Gabrielsson and Lindström, 2001; Gomez and Danuser, 2007; Nielzén and Cesarec, 1982). More than reflecting emotions, a central application of generative music, and in particular AAC, is to develop technology for building innovative intelligent systems that can not only monitor emotional states in the listener, but also induce specific affective states through music, automatically and adaptively.

This paper presents a listener evaluation of a prototype AAC system which targeted affective states by means of an underlying matrix of musical features derived by literature review. In this prototype we evaluate a system incorporating a range of musical features with known affective correlates; *tempo, mode, pitch range, timbre*, and *amplitude envelope*. Specific variations in each of these musical features are exploited in a the generative ruleset of an AAC system to imply different affective states in newly generated music samples, according to a 3x3 Cartesian grid across a 2-D affective space based on the circumplex model (Russell, 1980). A tri-stage listener evaluation was employed to inform two levels of subsequent adjustment to the feature mappings wherein the size and spread of

affective responses was gradually increased by deliberate manipulation of the combinations of musical features in the affective mappings, until a broad spectrum of emotional responses could be achieved.

1.1   Music and Emotion

Music has been shown to be capable of *emotional contagion* (Egermann and McAdams, 2013), and inducing physical responses on a conscious, and unconscious level (Grewe et al., 2007, 2005). A number of models have been adopted to measure affective responses to music. The circumplex (2-Dimensional) model is common, and is the model adopted in this experiment. In this model, *valence* represents positivity of the affective state, as plotted on the horizontal axis of a 2-Dimensional space, and *arousal* represents the intensity of the state, plotted on the vertical axis (Russell, 1980). Musicologists have tried to correlate affective responses with specific musical features, though discrete correlations that might be readily implemented in a computer aided composition system are rare. Many correlations between musical features and emotional descriptors exhibit overlap. For example, slower music is typically associated with lower *arousal* and *valence* — slow music might correlate with perceptual descriptors like *sad* or *tired* (Javela et al., 2008; Jefferies et al., 2008). A good deal of further work remains in uncovering the precise amount of overlap amongst these correlates, and the impact of an initial affective state on the exhibited response to the musical feature correlates.

1.2   Affectively-Driven Algorithmic Composition

Algorithmic composition, and the large variety of techniques for computer automation of algorithmic composition processes, are well documented in literature. Surveys of expressive computer performance systems such as that carried out by (Kirke and Miranda, 2009) also provide a thorough overview of the extensive work carried out in the area of emotionally targeted computer aided music performance. Rowe (Rowe, 1992) describes three distinct approaches in generic algorithmic composition systems: generative, sequenced, or transformative. Sequenced systems make use of pre-composed musical passages, which are subsequently ordered according to the algorithm. Generative systems create new musical passages according to particular rulesets (often, the selective filtering of random data). Transformative systems, the type evaluated in this paper, take existing musical passages as their source material and derive new sequences according to various functions which might be applied to this source material (for example, a basic transformation might be to reverse the notes of a musical sequence - commonly referred to as a retrograde transformation).

In the case of the experiments documented here, the intention ultimately is to create a system for automated generation of new music which can both reflect and induce a change in the listener's affective state. The distinction between perceived and induced affective state is important: the affective state of the listener must *actually* change in response to the musical stimulus in order for the state to be considered induced. This issue is well documented in music psychology (Gabrielsson, 2002; Kallinen and Ravaja, 2006; Scherer, 2004)with reference to traditional affective evaluations of musical compositions and performances, but with AAC, reactive, feedback-driven systems might induce affective states in the listener according to their own preferences and subsequent physiological responses. This distinction can be summarised as the difference between understanding a composer's intention (perceived emotion), and actually feeling a change (induced emotion) (Gabrielsson, 2002; Marin and Bhattacharya, 2010; Scherer, 2004).   For example, listeners in negative states of mind often find music which is intended to 'communicate' sadness to induce an increase in valence and arousal: sad listeners find sad music uplifting or enjoyable (Vuoskoski et al., 2012; Vuoskoski and Eerola, 2012). Thus, this is an area where the use of brain-computer interfacing or biophysical monitoring as a means of control for such systems could be extremely valuable to the field, if for example affective correlates of induced emotion could be accurately determined and subsequently used as a control signal for determining specific feature sets for AAC based music generation. In the first instance, however, this work documents progress in generating music with correctly perceived emotional content (i.e., music which is intended to convey a particular emotion, which is corroborated by listeners perception of the intended emotion).

2.   EMOTIONALLY INTELLIGENT MUSIC GENERATION SYSTEM

A series of affective mappings (musical features with emotional responses) was drawn from literature – see (D. Williams et al., 2014)for the full literature review – and implemented in an artificial intelligence driven AAC system. Initially a single musical feature, *rhythmic density*, was evaluated (Williams et al., 2015). This feature can contribute to perceived tempo (which, as mentioned above, has been suggested to be well correlated with affective arousal) as well as other sub-features (for example articulations like staccato or legato performance). This system was subsequently expanded to include a larger matrix of musical features; *tempo, mode, pitch range, timbre*, and

amplitude envelope, which could be used as affective correlates, creating an affective mapping for a fuller AAC system

These mappings utilize a time series of varying proportions of these musical features, intended to evoke particular affective states on the 2-Dimensional affective space. This prototype system uses these musical feature mappings to inform the generation of new music as a piano score, aiming for a full range of affective responses across the 2-D space. Seed material is generated by means of a 16-channel feed-forward artificial neural network (ANN). ANNs have previously been used in algorithmic music systems in (Bresin, 1998; Bresin and Friberg, 2011; Purwins et al., 2000). The ANN in this case is trained on 12 bars of piano music in C major at 120 bpm, as shown in Figure 1. Music is input as MIDI data and output in the same format. The ANN hidden layers are trained according to the weighting of each value in Table 1 from the seed material (off line). There is one layer for each half step in an octave (16 notes, rather than the 12 on a piano, i.e., we allow for both Bflat and Csharp etc), $a(n) \rightarrow pw(n)$ where p is the input vector, a is the output vector, and w is the weighting for each input as determined by the transformation/activation function (essentially, transformation f creates weighting w which then acts on input p to create output a). For each epoch:

- Initialize the neural network, set weights w_{ij} for each $_jth$ node for each area of the 2-dimensional emotion space
- Input neurons x_{ij} from target set of emotional correlates (musical features from matrix)
- Apply transformation/activation function f, $f(x)$
- Change weights w_{ij} of nodes in hidden layers
- Output layer (generate score for performance)

This approach has been used for many supervised learning applications, for example (Carpenter et al., 1992).

Fig. 1. Twelve bar source material used to train the ANN in the AAC pilot system, from Mozart's sonata K545. Note the material is in C Major (no key signature) and contains a variety of rhythmic patterns including a triplet and extended sequences of 8th/16th notes.

The seed pool of material is then further transformed according to the affective mapping of five musical sub-features: *tempo, mode, pitch range, timbre,* and *amplitude envelope*. Specific variations in each of these musical features are used to imply different affective states in the generated material according to a 3x3 Cartesian grid across the 2-D space, as shown in Figure 2, ranging from low valence and low arousal in the bottom left of the space, to high valence and high arousal in the top right of the space. Tonality is manipulated by selecting from a variety of scales and pitch quantizing the pitch class accordingly (*harmonic minor, chromatic, major, pentatonic*) and the generator than transposes that scale type so that it can cover all keys. Thus there is no additional requirement to list modes separately because as far as pitch class content is concerned they are identical to the parent scale. In other words, the only

difference between *C Maj* and *G mixolydian* is which note the scale starts on, (which is determined by the incoming MIDI notes), as the pitch class content is otherwise identical. *Tempo* ranges from 90,120,150 bpm (but these can be linearly interpolated). This is carried out as a number of initial messages sent to the pitch classes to generate note output. *Envelope* is a simple MIDI parameter amounting to sustain on/off and the generation of the legato/staccato effect, respectively. *Pitch spread* is a range between 2 and 6 octaves across the piano keyboard, as with tempo, these values can be linearly interpolated on the fly by the system when it is in generate mode. *Timbre* variation is simplistically implemented, we assume that brighter, harder timbres as created by more intense playing on the piano are mapped to the top 42 MIDI cc messages for 'velocity', with softer, gentler timbres in the bottom 42 MIDI cc messages for the same control value. As with pitch and tempo these values can be linearly interpolated on-the-fly when the system is in generate mode, and the ranges for all of these can be over-written when the ANN is offline.

arousal

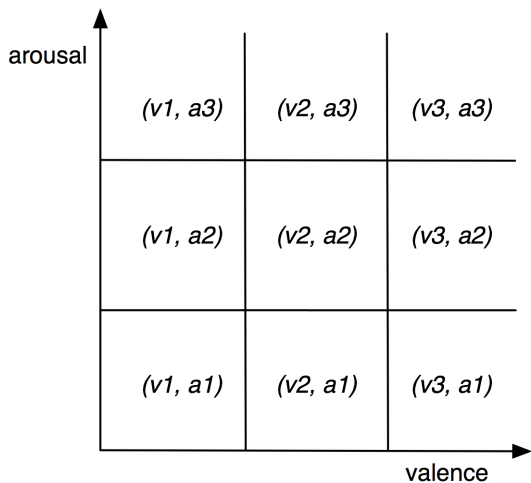| (v1, a3) | (v2, a3) | (v3, a3) |
| (v1, a2) | (v2, a2) | (v3, a2) |
| (v1, a1) | (v2, a1) | (v3, a1) |

valence

Fig. 2. 2-Dimensional affective space used for targeted stimulus generation, showing Cartesian co-ordinates in arousal (emotional intensity) and valence (positivity of emotion), after the circumplex model of affect (Russell, 1980).

The generated musical samples are performed by a synthesized piano voice. Co-ordinates with higher arousal generally include a larger pitch spread (range of notes), faster tempo, and harder timbres, whilst co-ordinates with higher valence generally utilize a major key. In this system, a Cartesian coordinate of (*valence [1:3]*, *arousal [1:3]*) is used to refer to a given combination of the five musical features, which are shown in Table 1. Thus a co-ordinate of (1, 1) would refer to low *valence* and low *arousal*, the lowest corner to the left of the space. This co-ordinate would force the transformation algorithm to create stimuli incorporating a slow tempo, a minor key, a *soft* timbre (on a piano, the timbre of the performance can be manipulated using dynamics markings, where perceptually *harder* sounds are achieved with louder performance dynamics), an amplitude envelope with considerable legato, and a spread of pitch values, which are comparatively lower than those of the rest of the generated pool. An example is given in Figure 3.

Fig. 3. Twelve bars of a generated musical figure with a target affective co-ordinate of (v1, a1), low valence and low arousal. Note that the musical features include a lower pitch spread (generated material is now in the bass clef) but retain some of the features of the training material (such as the single instance of a triplet, this time in bar 2) and instances of ¼ note rests as well as a mixture of mainly 1/8th and 1/16th note rhythms. Note also that although the key signature is empty, a number of accidentals, particularly a recurring b and e flat, imply the key of G minor (though the occasional F sharp makes this somewhat ambiguous).

The generic feature set used for each affective co-ordinate by the AAC system is shown in Table 1.

Table 1 Stimulus set showing generic musical feature matrix derived by literature review

Valence, Arousal	Timbre	Key	Pitch spread	Tempo	Envelope
Low, low (1,1)	Soft	Minor	Low	Slow	Legato
Medium, low (2,1)	Soft	Chromatic	Medium	Slow	Legato
High, low (3,1)	Soft	Major	High	Slow	Legato
Low, medium (1,2)	Medium	Minor	Low	Medium	None
Medium, medium (2,2)	Medium	Chromatic	Medium	Medium	None
High, medium (3,2)	Medium	Major	High	Medium	None
Low, high (1,3)	Hard	Minor	Low	Fast	Staccato
Medium, high (2,3)	Hard	Chromatic	Medium	Fast	Staccato
High, high (3,3)	Hard	Major	High	Fast	Staccato

2.1   Aim

A target spread of stimuli covering the maximum possible amount of the 2-D affective space was drawn from (Eerola and Vuoskoski, 2010), in which a large dataset of excerpts from film scores spanning a range of timbres and styles was rated by 116 participants across a variety of emotional descriptors, specifically including valence and energy, where energy can be taken as synonymous with arousal as in the precedent of other work examining musical arousal (Den Brinker et al., 2012; Ilie and Thompson, 2006), and orthogonal from valence (Shapiro and MacInnis, 2002). A subset of this dataset was drawn by means of principle component analysis such that the selected excerpts spanned the broadest possible range of the 2-D emotional space, as shown in Figure 4, which conceptually represents the best possible spread of affect in a real-world (i.e., not generated by means of artificial intelligence) stimulus set. However, 2-D emotion spaces can be problematic in mapping some musical emotions (for example, both *anger* and *fear* would be considered low valence and high arousal), and as such 3-D models including dominance have been suggested (Mehrabian, 1996) and may be useful in further work.
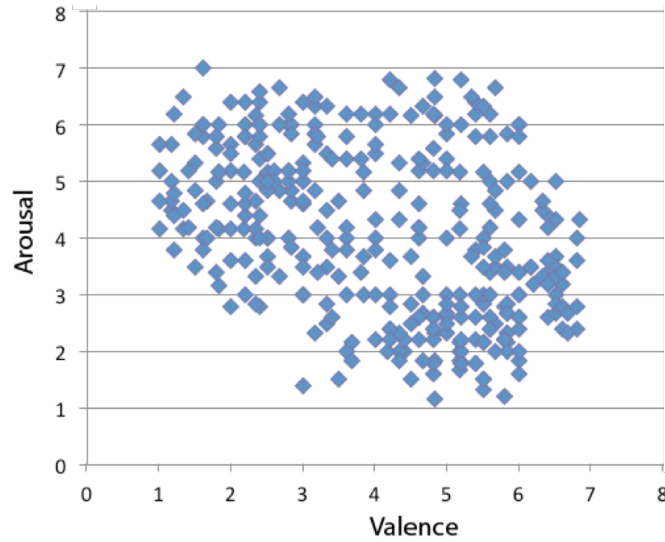
Fig. 4. Scatter plot showing target stimulus set drawn by PCA from (Eerola and Vuoskoski, 2010), in order to achieve maximal distribution across a 2-Dimensional affective space, comprising valence on the horizontal axis and arousal on the vertical axis, both rated on a 9-point scale (from 0-8).

The goal of these experiments was to derive a ruleset for computer generated music which can cover a comparable spread of emotions as the existing real-world sample set shown in Figure 4. There is no specific requirement for the system to have 'pleasant' or 'unpleasant' performances by default, but rather to establish that the system can create performances which are capable of targeting the fullest possible range of affective responses. The reader should note that it is not the intention of this paper to compare the generated music to the real-world sample set (which would require a separate experiment with direct comparison), but rather to compare the possible spread of affect and the calibration of the generator by means of adjusting the musical feature correlations involved.

3. EXPERIMENT

Three stimulus iterations, each comprising 99 musical excerpts generated by the prototype system documented above, were rated by a listening panel of 36 participants. Each participant had some experience of critical listening as part of undergraduate level study in music or music technology. 13 participants were female, 23 participants were male, and all participants were between 20 and 34 years old. 7 additional participants were rejected from the listening panel on the basis of a pre-experiment Short Test of Musical Preferences (STOMP) questionnaire, (Rentfrow and Gosling, 2003), which was conducted in order to exclude listeners with extreme preferences for particular genres (specifically excluding listeners who rated below the 50th percentile for enjoyment of western classical music, referred to as the reflective and complex domain in the STOMP classification).

3.1 Ethical Approval

Ethical approval to undertake this experiment was applied for on 4th April 2015 and granted by the research ethics sub-committee of the faculty of arts, Plymouth university, on 6th June 2015, after revisions to the information sheet given to participants had been completed. All participants were informed that this would be the case and that their data might be subject to repeated or new analysis as a secondary dataset by a third party. Participants were free to opt-out or leave the experiment at any point.

3.2 Method

Stimuli for each Cartesian co-ordinate were generated to make a complete stimulus set comprising 99 twenty-second passages of music (11 for each co-ordinate) for each iteration of the experiment.

The experiments were conducted via a customised graphical user interface, shown in Figure 5, which combined elements of the self assessment manikin (SAM), in order to provide pictorial analogies for valence and arousal (Bradley and Lang, 1994) evenly across a 2-D space. Playback was conducted via circumaural headphones in a quiet room with a dry acoustic. Participants were allowed to adjust the volume of the playback level according to their own preference at any point during the task via the GUI. The order of stimulus presentation was randomised for each participant. Each listener evaluated the complete stimulus set.

Participants were instructed to listen to each piano sample, and rate the emotional content on the 2-D space using the GUI, as shown in Figure 5. Participants could return to any ratings or repeat stimuli at any time. A complete session would typically take between 35 to 40 minutes.
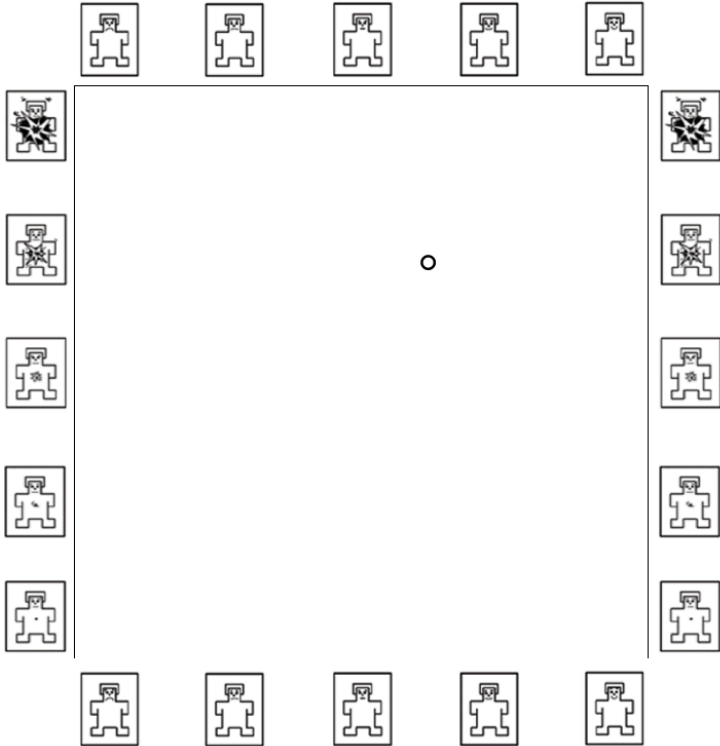

Fig. 5. The GUI that participants used to evaluate each stimulus. Versions of the self-assessment manikin (SAM) are placed on each horizontal and vertical axis, illustrating valence on the horizontal axis and arousal on the vertical axis, with even spacing between each emoticon. A hidden grid divides each rating into 9x9 possible responses (1-9 valence, 1-9 arousal). Participants are able to adjust the playback volume independently according to their own preference, and repeat or return to any stimuli to re-rate them as they choose.

3.3 Results

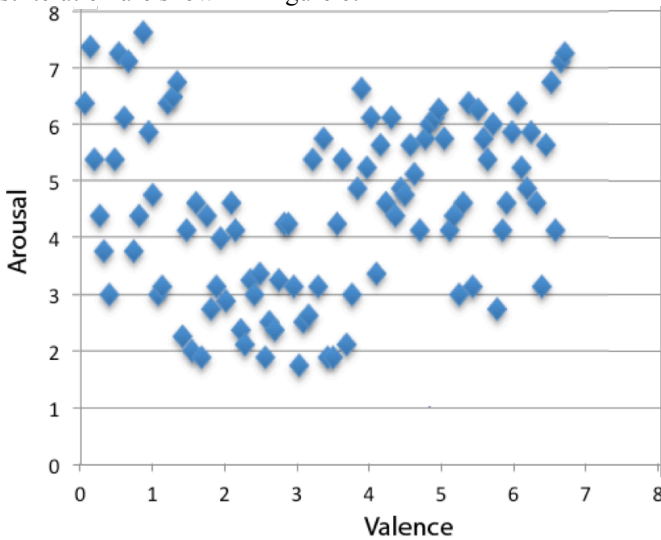Listener responses to the first iteration are shown in Figure 6.

Fig. 6. Scatter plot showing the mean responses to each of the stimuli after first test

The mean standard deviation across listener responses (the spread of ratings by listeners to the same stimulus) to the first iteration was 1.426. The spacing of the stimulus set showed that the system could create music across a wide range of affective responses, but with notable clustering towards the top left of the space (1, 3, or angry/afraid), through the lower middle of the space, and towards the top right of the space. Particular absences can be seen towards the upper middle and the lower right of the space ((3, 1), or calm).

The combination of musical features was adjusted to reflect this, such that stimuli which were intended to have an affective co-ordinate of (3,1) used a slower tempo, and stimuli that were intended to have an affective co-ordinate of (2,3) were generated with faster tempo and more spectrally dense (harder) timbres. A new set of 99 stimuli was then generated based on these adjustments and evaluated by the same group of listeners one week after the first iteration was conducted. Responses to the second iteration are shown in Figure 7.
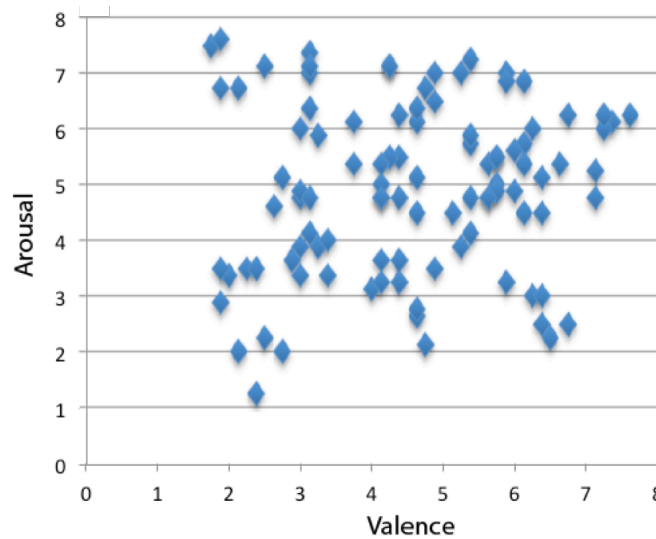

Fig. 7. Scatter plot showing the mean responses to each of the stimuli after the second test

The mean standard deviation across listener responses to the second iteration was 1.035. The spacing of responses to the second iteration of stimuli suggested that a wider affective range had been achieved, with notable improvement in the lower right corner of the space (1, 3, or calm), and in the upper middle of the space (2, 3). Some of the clustering in the top left of the space was improved upon, with some notable absences in the middle left (1, 2), and lower middle (2, 1) areas of the space.

The musical feature set was selectively adjusted to force stimuli in the lower middle of the space to be quantized to a pentatonic mode, with stimuli in the middle left of the space to remain in a minor mode with a softer timbre and a slightly slower tempo than the previous ruleset specified. A third set of 99 stimuli was then generated and evaluated by the same group of listeners, two weeks after the first iteration had been conducted. Responses to the third iteration are shown in Figure 8.
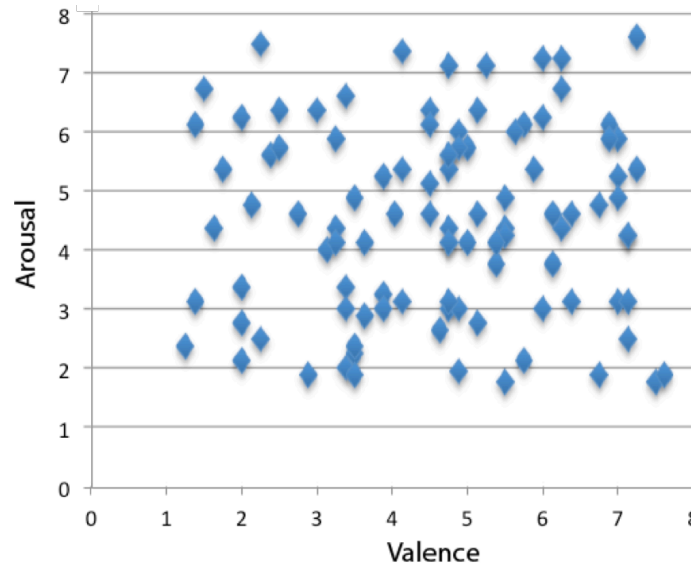
Fig. 8. Scatter plot showing the mean responses to each of the stimuli after the third test

The mean standard deviation across listener responses to the third iteration was 0.834, suggesting that there was more inter-participant agreement across the spread of listener responses to this stimulus set than in either of the previously generated iterations. The spread of listener responses across all iterations was then examined both visually and statistically. Whilst the difference may not be entirely obvious visually, there was a marked increase in variance and standard deviation from the 1st to the 3rd iteration, shown in Table 2. These measures of spread summarize the responses to each iteration to show how distributed the perceived emotional values were for each iteration, and how much they differ from the mean responses.

Table 2 Measures of spread across the stimulus set iterations

Measure	1st iteration	2nd iteration	3rd iteration
Variance	2.343	2.426	2.879
Standard Deviation	1.530	1.557	1.697
1st Quartile	3.625	3.125	3.375
3rd Quartile	6.125	5.750	6.000
Inter Quartile Range	2.500	2.625	2.625

The inter quartile range across the spread of responses showed an improvement from the 1st to 2nd iteration, but no improvement from the 2nd to the 3rd iteration. This makes concluding an improvement in the spread more challenging, but the visual observation of the distribution, combined with the consistent increase in standard deviation and variance combined to suggest that the spread of listener responses to the third iteration of stimuli could still be considered a notable improvement in spread over both previous iterations with limited clustering occurring around the centre of the space (arguably the perceptually 'neutral' stimuli). This is perhaps because 'neutral' music is not trivial for listeners to quantify, and indeed it may be the case that in future, generation of 'neutral' music is thus not readily generatable by AAC. The final spread of ranges from (1,2) to (7,1) on the bottom of the space, from (1,7) to (7,7) on the top of space, represents a marked improvement over the real-world target which ranged from (1,3) to (7,3) on the bottom of the space, and (1,7) to (5,6) on the top of the space. Of the 81 possible co-ordinates that might be indexed by a range of 1-9 in valence and 1-9 in arousal, the target real-world stimuli offered access to approximately 19/81 co-ordinates – around 23% – with the generated stimuli offering access to approximately 33/81 possible co-ordinates – around 40% – representing a 17% improvement in overall coverage across the conceptual emotion space, and a 42.5% relative improvement on the spread of the target ratings. Neither stimulus set was capable of accessing the most extreme co-ordinates in any of the four corners of the conceptual space.

4. CONCLUSIONS

Listener evaluation of a prototype AAC system confirmed that affective targeting via novel musical material is possible if the material is generated by means of an underlying matrix of musical features with documented emotional correlations. A good spread of emotional responses was achieved, broadly comparable to the spread in a real-world stimulus set, but the real-world stimulus set was not directly evaluated against the generated stimulus set, and thus the generator should be considered as conceptually distinct from the real-world stimulus set, a direct comparison would require further experiment. The basic technique for generating music described here is not novel in and of itself. Many other successful implementations of different types of probabilistic analysis and generation exist, including Markov models (Casey, 2001; Visell, 2004) and recurrent neural networks (Bown and Lexer, 2006), amongst others. However, the combination of generative AI techniques with affective feature mapping does offer a novel way of creating and interacting with emotionally expressive music, with enormous potential in the context of creating opportunities for musical communication, for example with a virtual co-performer (Williamon and Davidson, 2002). The field of AAC is in its infancy and therefore there are no major precedents for evaluation of such a system. Methods for evaluating the utility of music creation systems, particularly algorithmic composition are the subject of much debate, including in some cases debate as to the question of authorship of music created by such systems (Dahlstedt, 2001; Eigenfeldt, 2011). These concerns remain important in cases where this type of system would be used for creative ends, though they are less important for applications such as the aforementioned music therapy or non-linear soundtrack generation for video gaming. Nevertheless, combining these techniques with deliberate affective targeting, in response to a real-time musical performance, does compound these issues. Therefore, beyond perceptual calibration of the musical feature emotion space mapping, and other issues for practical implementation in the generative system, there remains a significant amount of further work to address in devising practical and robust evaluation methodologies for AAC systems. Should such a system be adaptable to real-time control by means of biophysiological estimates of affective state, a feedback driven AAC system could be created for continuous monitoring and induction of target affective states in the listener (e.g., for therapeutic means) in the future. In this work the particular combination of these features has been explored and adjusted in response to listener evaluation to successfully exploit a large portion of a 2-Dimensional emotion space, but the complex nature of the inter-relationship between these musical features, and the subsequent affective responses that might be the subject thereof, remains an area of considerable further work.

5. ACKNOWLEDGMENTS

REFERENCES

Bown, O., Lexer, S., 2006. Continuous-time recurrent neural networks for generative and interactive musical performance, in: Applications of Evolutionary Computing. Springer, pp. 652–663.

Bradley, M.M., Lang, P.J., 1994. Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psychiatry 25, 49–59.

Bresin, R., 1998. Artificial neural networks based models for automatic performance of musical scores. J. New Music Res. 27, 239–270.

Bresin, R., Friberg, A., 2011. Emotion rendering in music: Range and characteristic values of seven musical variables. Cortex 47, 1068–1081. doi:10.1016/j.cortex.2011.05.009

Brown, E., Cairns, P., 2004. A grounded investigation of game immersion, in: CHI'04 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 1297–1300.

Carpenter, G.A., Grossberg, S., others, 1992. A self-organizing neural network for supervised learning, recognition, and prediction. IEEE Commun. Mag. 30, 38–49.

Casey, M., 2001. General sound classification and similarity in MPEG-7. Organised Sound 6. doi:10.1017/S1355771801002126

Dahlstedt, P., 2001. A MutaSynth in parameter space: interactive composition through evolution. Organised Sound 6. doi:10.1017/S1355771801002084

Den Brinker, B., Van Dinther, R., Skowronek, J., 2012. Expressed music mood classification compared with valence and arousal ratings. EURASIP J. Audio Speech Music Process. 2012, 1–14.

Eaton, J., Williams, D., Miranda, E., 2014. Affective Jukebox: A confirmatory study of EEG emotional correlates in response to musical stimuli.

Eerola, T., Vuoskoski, J.K., 2010. A comparison of the discrete and dimensional models of emotion in music. Psychol. Music 39, 18–49. doi:10.1177/0305735610362821

Egermann, H., McAdams, S., 2013. Empathy and emotional contagion as a link between recognized and felt emotions in music listening. Music Percept. Interdiscip. J. 31, 139–156.

Eigenfeldt, A., 2011. Real-time Composition as Performance Ecosystem. Organised Sound 16, 145–153. doi:10.1017/S1355771811000094

Gabrielsson, A., 2002. Emotion perceived and emotion felt: Same or different? Music. Sci. 5, 123–147.

Gabrielsson, A., Lindström, E., 2001. The influence of musical structure on emotional expression., in: P. N. Juslin J. A. Sloboda (Ed.), Music and Emotion: Theory and Research, Series in Affective Science. Oxford University Press, New York, NY, US, pp. 223–248.

Gomez, P., Danuser, B., 2007. Relationships between musical structure and psychophysiological measures of emotion. Emotion 7, 377.

Grewe, O., Nagel, F., Kopiez, R., Altenmüller, E., 2007. Emotions over time: Synchronicity and development of subjective, physiological, and facial affective reactions to music. Emotion 7, 774–788. doi:10.1037/1528-3542.7.4.774

Grewe, O., Nagel, F., Kopiez, R., Altenmüller, E., 2005. How Does Music Arouse "Chills"? Ann. N. Y. Acad. Sci. 1060, 446–449.

Grimshaw, M., Lindley, C.A., Nacke, L., 2008. Sound and immersion in the first-person shooter: mixed measurement of the player's sonic experience, in: Proceedings of Audio Mostly Conference. pp. 1–7.

Ilie, G., Thompson, W.F., 2006. A comparison of acoustic cues in music and speech for three dimensions of affect. Music Percept. 23, 319–330.

Javela, J.J., Mercadillo, R.E., Ramirez, J.M., 2008. Anger and associated experiences of sadness, fear, valence, arousal, and dominance evoked by visual scenes. Psychol. Rep. 103, 663–681.

Jefferies, L.N., Smilek, D., Eich, E., Enns, J.T., 2008. Emotional valence and arousal interact in attentional control. Psychol. Sci. 19, 290–295.

Kallinen, K., Ravaja, N., 2006. Emotion perceived and emotion felt: Same and different. Music. Sci. 10, 191–213. doi:10.1177/102986490601000203

Kirke, A., Miranda, E.R., 2009. A survey of computer systems for expressive music performance. ACM Comput. Surv. 42, 1–41. doi:10.1145/1592451.1592454

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.

Mehrabian, A., 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Curr. Psychol. 14, 261–292.

Nielzén, S., Cesarec, Z., 1982. Emotional experience of music as a function of musical structure. Psychol. Music.

Purwins, H., Blankertz, B., Obermayer, K., 2000. Computing auditory perception. Organised Sound 5. doi:10.1017/S1355771800005069

Rentfrow, P.J., Gosling, S.D., 2003. The do re mi's of everyday life: the structure and personality correlates of music preferences. J. Pers. Soc. Psychol. 84, 1236.

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.

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.

Shapiro, S., MacInnis, D.J., 2002. Understanding program-induced mood effects: Decoupling arousal from valence. J. Advert. 31, 15–26.

Visell, Y., 2004. Spontaneous organisation, pattern models, and music. Organised Sound 9. doi:10.1017/S1355771804000238

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.

Williamon, A., Davidson, J.W., 2002. Exploring co-performer communication. Music. Sci. 6, 53–72.

Williams, D., Kirke, A., Miranda, E., Daly, I., Hallowell, J., Hwang, F., Roesch, E., Nasuto, S.J., 2014. Evaluating perceptual separation in a pilot system for affective composition., in: Proceedings of the 40th International Computer Music Conference. ICMC, Athens, Greece.

Williams, D., Kirke, A., Miranda, E., Daly, I., Hallowell, J., Weaver, J., Malik, A., Roesch, E., Hwang, F., Nasuto, S., 2015. Investigating Perceived Emotional Correlates of Rhythmic Density in Algorithmic Music Composition. ACM Trans. Appl. Percept. In press.

Williams, D., Kirke, A., Miranda, E.R., Roesch, E., Daly, I., Nasuto, S., 2014. Investigating affect in algorithmic composition systems. Psychol. Music. doi:10.1177/0305735614543282