EVALUATING PERCEPTUAL SEPARATION IN A PILOT SYSTEM FOR AFFECTIVE COMPOSITION

Duncan Williams, Alexis Kirke, Eduardo Miranda

Interdisciplinary Centre for Computer Music research,
Plymouth University, Devon, PL4 8AA, UK
duncan.williams@plymouth.ac.uk

Ian Daly, Etienne B. Roesch, James C. E. Weaver, Slawomir J. Nasuto

Brain Embodiments Laboratory, University of Reading,
Whiteknights, Reading, RG6 6AH, UK

ABSTRACT

Research evaluating perceptual responses to music has identified many structural features as correlates that might be incorporated in computer music systems for affectively charged algorithmic composition and/or expressive music performance. In order to investigate the possible integration of isolated musical features to such a system, a discrete feature known to correlate some with emotional responses – rhythmic density – was selected from a literature review and incorporated into a prototype system. This system produces variation in rhythm density via a transformative process. A stimulus set created using this system was then subjected to a perceptual evaluation. Pairwise comparisons were used to scale differences between 48 stimuli. Listener responses were analysed with Multidimensional scaling (MDS). The 2-Dimensional solution was then rotated to place the stimuli with the largest range of variation across the horizontal plane. Stimuli with variation in rhythmic density were placed further from the source material than stimuli that were generated by random permutation. This, combined with the striking similarity between the MDS scaling and that of the 2-dimensional emotional model used by some affective algorithmic composition systems, suggests that isolated musical feature manipulation can now be used to parametrically control affectively charged automated composition in a larger system.

1. INTRODUCTION

Computer music systems for algorithmic composition can use both musical feature-sets and specifications for isolated musical features as input rules. Whilst many such systems exist, research documenting the precise affective correlation of isolated musical features is sparse. In the future, affective correlations to these musical features might be exploited by systems for emotionally-driven algorithmic composition. However, perceptual evaluations of discrete musical features in the context of these affective correlations are not readily available. This paper presents work towards this goal by implementing an isolated musical feature in a prototype system and subjecting the generated output to a perceptual evaluation.

When considering the selection of an appropriate musical feature to implement, previous research discussing affective *performance* algorithms confirmed that feature choice is a complex issue [1]. Therefore, a survey of affective responses to musical features in literature was carried out in order to determine likely correlates for affective algorithmic composition. Interested readers can find more exhaustive reviews on the link between music and emotion in [2] and the recent special issue in Musciae Scientiae [3].

1.1 Musical features as perceptual vectors

Whilst some musical features have well-defined acoustic cues, others have more complicated, even overlapping cues. Pitch, for example, is well correlated acoustically with fundamental frequency, whilst tremolo is well correlated with amplitude envelope. Meter, on the other hand, which has been found to be correlated with some emotions by Kratus [4], correlates with both frequency and time-derived acoustic cues [5] as a combination of duration, accent, and repetition. Therefore an awareness of listeners' methods for perceiving such features, and any hierarchical interaction between such features becomes important when selecting an isolated musical feature for experimentation.

A literature review of existing systems for affectively driven algorithmic composition suggested that modality, rhythm, and melody had been most commonly implemented, with 29, 29, and 28 instances respectively in the literature [6]. Other major features that had been implemented by systems surveyed in the literature included timbre, dynamics, tempo, and articulation. Of the two most popular features, 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 systems for affective algorithmic composition. However, for the purposes of this prototype system and its evaluation, rhythm would be a difficult selection as an isolated feature for perceptual evaluation due to the complex interaction of many of the sub-features and contributory acoustic cues involved. Therefore, the most common sub-feature of rhythm was chosen for the prototype system in this experiment, in order to minimize unwanted interaction from other musical features.

2. PROTOTYPE DESIGN

Rowe [7] describes three methodological approaches to algorithmic composition: generative, sequenced, or transformative. Transformative systems use existing material as the source – one or more transformations are applied to the input material in order to yield related material at the output stage. A simple inversion might be considered an example of a transformative process. Given the successful implementation of previous transformative systems (see for example, the *Experiments in Musical Intelligence* work of Cope [8]–[10]), the prototype system was designed to use a transformation algorithm in order to manipulate the selected, isolated musical feature (rhythmic density – a temporal aspect of music derived from pulses or beats, tempo, and meter). Transformative systems have the advantage over generative systems of an 'in-built' limiting rule-set, established by the seed material. With a transformative system, for example, there is no necessity to specify a large body of additional structural rules outside of those affected by the chosen feature. If such a system can be shown to achieve perceptual variation with a limited musical rule set, then in the future it should be adaptable to generative operation by the addition of appropriate structural rules (that need not be based on musical feature selection or perceptual correlation). This would also facilitate work towards a larger system for affective algorithmic composition based on the selective manipulation of a broader range of musical features with underlying emotional correlations.

2.1 Transformative algorithm

The prototype system was developed using OpenMusic [11] and Common Lisp. The system functions offline and currently works with monophonic data only. The system has three phases; a learning phase, a transformation phase, and a generation phase. At the learning phase, the system takes a seed input and separates the musical structure into measures, deriving a two-order transition matrix of pitch and rhythm tree information (a hierarchical list representing rhythmic structures with probability values for the transitions between these structures). These values are stored as an array. A statistical analysis of rhythmic density is then carried out on the array by searching for the number of pulses in each measure via note onset and duration values. Each measure is then assigned a density value. At the transformation phase, the density value is used as an index in order to create new permutations from a Markov chain of pitch and rhythm tree information via the transition matrix. Permutations can be created solely from measures with high-density index values, low-density index values, or a combination (assuming that there is enough variation in the original seed material). The permutations are used by the generation phase to allow the output to be saved as a MIDI format file for subsequent editing and playback. A signal flow of the prototype system is shown in Figure 1. If successful, the prototype system could be expanded by increasing the *n*th-order of the Markov chain to include more complex transitions, other musical features and higher level musical structures, once the relevant perceptual correlations have been determined.

3. PERCEPTUAL EVALUATION

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 from listener perceptions of music [12]–[15]. Confidence in the model can be evaluated by statistical measures from the analysis in order to firstly determine the best-fit dimensionality for the model, and secondly to create a plot of the stimuli showing respective and relative similarities in the model. With all MDS analysis, dimensional labels cannot be established by this kind of evaluation. This experiment therefore represents the first of a two-stage validation of the prototype system – with the expectation that, if successful, a second stage will include a verbal elicitation experiment to provide labels for the scaled data. Stimuli should be presented to the listeners in the second experiment in the order that they are arranged in the best-fit perceptual space from this experiment, with the aim being to provide meaningful emotional (or at least perceptual) labels for the movement in each of the resulting dimensions.

3.1 Stimulus set generation/selection

Stimuli for the experiment were created using the prototype system and 4 seed inputs from a study evaluating affective responses and neurophysical correlations in electroencephalogram (EEG) to western classical music [16]. These seed inputs were selected with the partial intention of adapting a BCMI system to the control of an affective algorithmic composition system using EEG in future. Thus, seed material which had already been perceptually evaluated with EEG seemed to be a useful starting point. The sources from [16] were *Peter and the Wolf* (Prokofiev), *Brandenburg Concerto No. 5* (J.S. Bach), *Four Seasons: Spring* (Vivaldi), and *Adagio for Strings* (Barber). This seed material was edited to produce short excerpts of 30s in duration, in order to facilitate timely comparisons in the experiment itself, and reduce listener fatigue.

With MDS analysis, a minimum of 4 stimuli per dimension to be revealed in the final analysis is required. 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:

- 1-4: original material, edited in duration only
- 5-8: lower density rhythmic transformations applied to seed material
- 9-12: higher density rhythmic transformations applied to seed material
- 13-16: permutation only (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).

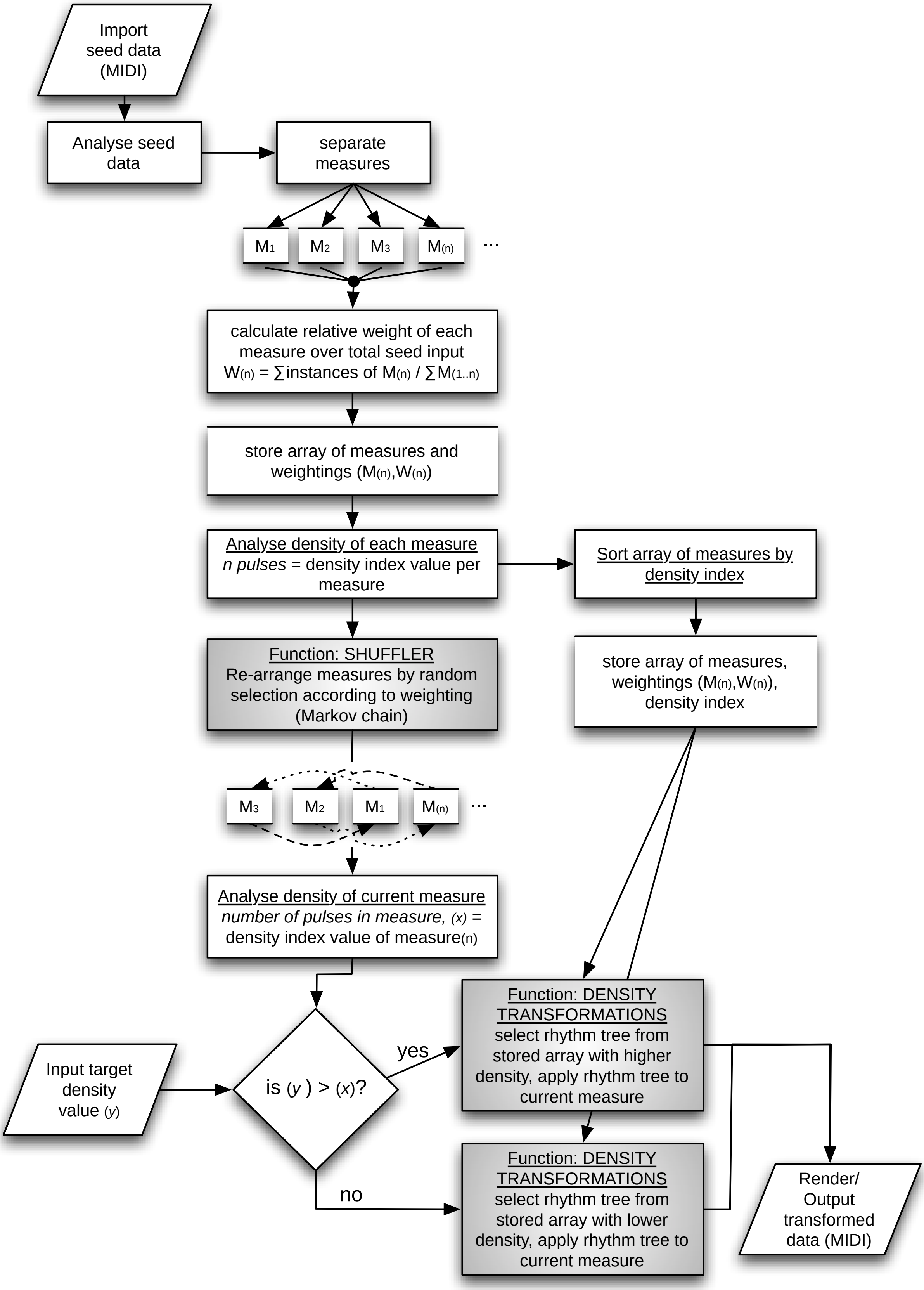

Figure 1. Signal flow of prototype system. Various permutations are possible, including the generation of a permuted set of measures using existing rhythm trees, a permuted set of measures with increased density (number of pulses extracted from other measures), and a permuted set of measures with decreased density.

Figure 2 shows an excerpt from the seed material before it has been separated by the system into measures. Figure 3 shows a lower density excerpt generated from the same seed by the prototype system. The seed in this case was an excerpt from J.S. Bach's *Brandenburg Concerto No. 5*, which mainly consists of 1/16th notes, with the exception of the material in the latter half of the sample. 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. The score itself is not optimised by the routine and could be edited by hand for ease of reading.

Figure 2. Excerpt of seed material, condensed to a monophonic piano arrangement of taken from J.S. Bach's *Brandenburg Concerto No. 5*

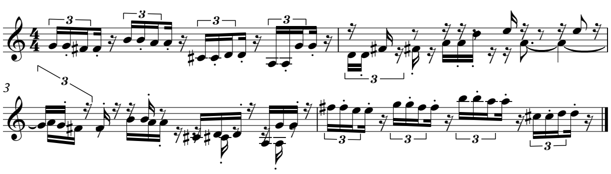

Figure 3. 'Lower density' excerpt created by markov permutation of measures from seed material, with a lower density index used as the basis for selection of rhythm trees. Note the algorithm has made use of triplets to emulate the pattern from the latter half of the seed material.

3.2 Experiment procedure and listening panel

Twenty two listeners participated in the experiment. 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 near-simultaneously (participants broadly began the experiment at the same time, in the same room) via 22 iterations of a Max/MSP graphical user interface on desktop computers. The same brand and model of circumaural headphones was used by all participants. Participants were allowed to adjust volume levels according to their own preference during a familiarization exercise. The familiarization exercise also allowed listeners to hear the full range of stimuli in a non-linear fashion before undertaking the main experiment.

In the main experiment, listeners were presented with 136 randomly ordered pairs of stimuli, split over two tests of approximately 35 minutes in duration. Listeners were asked to compare and rate the similarity between each pair on a hidden 100-point scale with end-points labeled 'not at all similar' and 'the same'.

4. RESULTS

Listener responses were collated to produce a dissimilarity matrix which was then subjected to an Individual Differences Scaling (INDSCAL) MDS analysis. The statistical 'measures-of-fit' determined by the analysis (dimensionality, RSQ or square of the correlation coefficient, and Kruskal stress) are shown in Table 1.

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

Table 1. 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).

As with any MDS analysis, increasing the number of dimensions will decrease the amount of stress on the solution, hence determining the optimum solution is not simply a matter of looking for the lowest stress. Hence, the statistical measures in Table 1 were then examined to determine the 'correct' dimensionality (the number of dimensions which best represented the perceived variation in the stimulus set). Criteria which can be used as indicators of statistical quality in such analysis include RSQ greater than 0.95 [17], stress greater than 0.20 and optimally as low as 0.05 [18], and a negligible improvement in RSQ at the next increase in dimensionality. Table 1 shows that RSQ was greater than 0.95 in all dimensionalities, suggesting that each gave 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 was below the threshold of <0.20 in all other solutions. Examination of a scree plot showing stress against dimensionality showed a significant knee (which can also be interpreted as an indicator of 'correct' dimensionality), at 2-Dimensions, shown in Figure 4. Together, these results strongly suggested a 2-D solution. The spread in a Shepard diagram at 2-Dimensions, as shown in Figure 5, was also examined, with a low spread in the data confirming a statistically good fit.

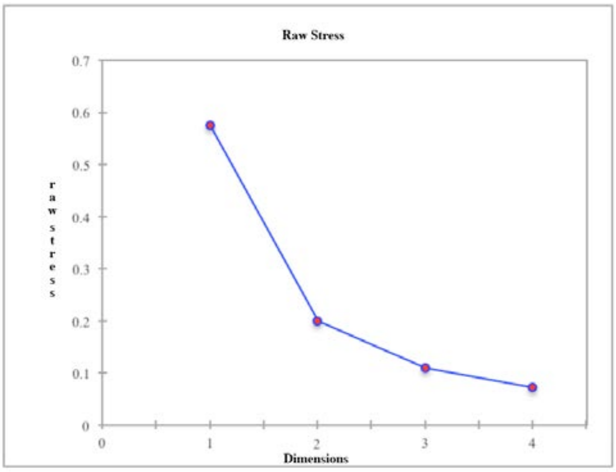

Figure 4. Scree plot showing a significant knee at 2-D, with stress at 0.200.

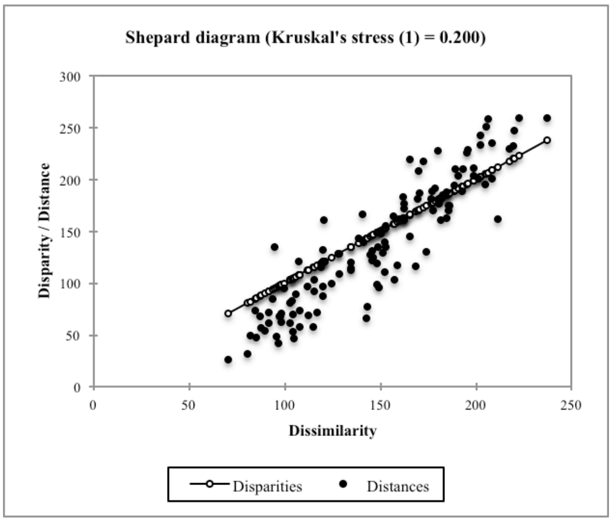

Figure 5. Shepard diagram showing a low spread between similarities and distances in the 2-D solution with stress at 0.200.

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

5. DISCUSSION

The permuted stimuli, as shown in Figure 6, are generally plotted closer to the respective seeds than the density transformations, suggesting that a permutation in overall musical structure has less perceptual significance to listeners than a variation in rhythmic density. This is somewhat surprising as it suggests that even when the output is modified significantly by this process of random permutation, the output retains more perceptual similarity to the seed material than the output generated by selectively and deliberately manipulating rhythmic density.

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 – a slow, sparse piece of music in comparison to the other seed sources. Similarly, 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 permutation remains the closest stimulus 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 this unexpected manner. 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, but 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.

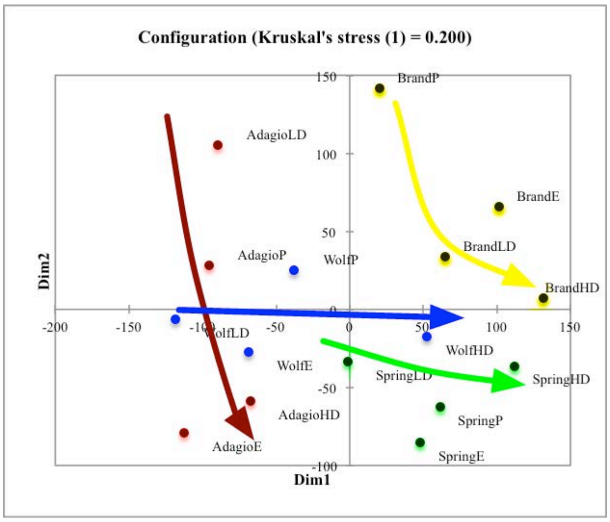

Figure 6. Perceptual space in 2-Dimensions after MDS INDSCAL analysis. Movement can be seen from low to high density stimuli. Coloured annotations show grouping of stimuli based on seed material. Stimuli appended -E are original edited seed excerpts. Stimuli appended -P are permutations with no intended 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 transformed stimuli are loosely grouped near to their seed material, with a general trend that low density transformations are found in the upper left of their seed group, and high density transformations in the lower right of their 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 [19], a 2-Dimensional emotional space with dimensions based on arousal and valence, which has been adapted to music and to affectively-charged algorithmic composition in some systems [20], [21]. Whilst such observations can only be casually drawn, Barber's *adagio* seems to be a 'sadder', more somber piece, whilst the Brandenburg concerto is faster, more lively, higher energy and subjectively 'happier', found at the opposite end of the 2-D space. This strongly suggests that isolated musical feature manipulation is compatible with this method of parameterizing affect in such systems, and that in the future, a larger system, incorporating several such isolated features as part of an affective control system, should be possible. A larger system would have the advantage of being able to generate affectively charged music automatically, and reactively, responding to the user's emotional state. However, as MDS analysis cannot reveal the names of dimensions given by this analysis, a subsidiary verbal elicitation experiment should now be undertaken before rhythmic density could be included in such a system. 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.

6. CONCLUSIONS

In order to determine whether isolated musical features could be used in a larger affective algorithmic composition 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 pairwise dissimilarity experiment.

The pairwise dissimilarity experiment concluded that listener responses could be plotted to a 2-Dimensional solution with reasonable statistical confidence. A subsequent verbal elicitation experiment could now be used to label these dimensions. Within the 2-Dimensional space, randomly permuted stimuli were found to be perceptually more similar to the seed material than stimuli created with deliberate variation in rhythmic density. This is a surprising finding and has implications for the incorporation of a larger range of isolated musical features in an affective algorithmic composition system.

The 2-Dimensional MDS scaling 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 contribute to a larger system for affectively charged algorithmic composition in the future.

Acknowledgments

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

7. REFERENCES

[1] A. Friberg, E. Schoonderwaldt, and A. Hedblad, "Perceptual ratings of musical parameters," *Gemessene Interpret.-Comput. Aufführungsanalyse Im Kreuzverhör Diszip. Mainz Schott*, 2011.

[2] K. R. Scherer, "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?," *J. New Music Res.*, vol. 33, no. 3, pp. 239–251, Sep. 2004.

[3] A. Lamont and T. Eerola, "Music and emotion: Themes and development," *Music. Sci.*, vol. 15, no. 2, pp. 139–145, Jul. 2011.

[4] J. Kratus, "A developmental study of children's interpretation of emotion in music," *Psychol. Music*, vol. 21, no. 1, pp. 3–19, 1993.

[5] E. E. Hannon, J. S. Snyder, T. Eerola, and C. L. Krumhansl, "The Role of Melodic and Temporal Cues in Perceiving Musical Meter.," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 30, no. 5, pp. 956–974, 2004.

[6] D. Williams, A. Kirke, E. R. Miranda, E. B. Roesch, and S. J. Nasuto, "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*, 2013.

[7] R. Rowe, *Interactive music systems: machine listening and composing*. MIT press, 1992.

[8] D. Cope and M. J. Mayer, *Experiments in musical intelligence*, vol. 12. AR Editions Madison, WI, 1996.

[9] D. Cope, "Computer modeling of musical intelligence in EMI," *Comput. Music J.*, pp. 69–83, 1992.

[10] D. Cope, "Experiments in musical intelligence (EMI): non-linear linguistic-based composition," *J. New Music Res.*, vol. 18, no. 1–2, pp. 117–139, 1989.

[11] J. Bresson, C. Agon, and G. Assayag, "Openmusic 5: A cross-platform release of the computer-assisted composition environment," in *10th Brazilian Symposium on Computer Music, Belo Horizonte, MG, Brésil*, 2005.

[12] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet, "Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts," *Cogn. Emot.*, vol. 19, no. 8, pp. 1113–1139, Dec. 2005.

[13] E. Bigand, S. Filipic, and P. Lalitte, "The time course of emotional responses to music," *Ann. N. Y. Acad. Sci.*, vol. 1060, no. 1, pp. 429–437, 2005.

[14] S. Vieillard, I. Peretz, N. Gosselin, S. Khalfa, L. Gagnon, and B. Bouchard, "Happy, sad, scary and peaceful musical excerpts for research on emotions," *Cogn. Emot.*, vol. 22, no. 4, pp. 720–752, 2008.

[15] T.-L. Wu and S.-K. Jeng, "Probabilistic estimation of a novel music emotion model," in *Advances in Multimedia Modeling*, Springer, 2008, pp. 487–497.

[16] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cogn. Emot.*, vol. 15, no. 4, pp. 487–500, 2001.

[17] B. Astill, "An investigation of social values in a senior secondary school milieu," in *In Proceedings of the Austalian Association for Research in Edu-*

cation Conference, University of Adelaide, Australia, 1994.

[18] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[19] J. A. Russell, "A circumplex model of affect.," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, p. 1161, 1980.

[20] I. Wallis, T. Ingalls, E. Campana, and J. Goodman, *A Rule-Based Generative Music System Controlled by Desired Valence and Arousal*. SMC, 2011.

[21] A. Kirke and E. Miranda, "Combining EEG Frontal Asymmetery Studies with Affective Algorithmic Composition and Expressive Performance Models," 2011.