Artificial Social Composition: A Multi-Agent System for Composing Music Performances by Emotional Communication

Alexis John Kirke and Eduardo Reck Miranda
Interdisciplinary Centre for Computer Music Research, University of Plymouth, UK
alexis.kirke@plymouth.ac.uk; eduardo.miranda@plymouth.ac.uk

Abstract

We present a multi-agent system based on an evolutionary model used in computational musicology – a Social Composition System. Agents are given an emotional state and an abstract performance model based on Director Musices - a computer system for expressive performance. As agents attempt to communicate their emotions, a simple composition process emerges based on the expressive performance between agents, the population's emotional profile, and the emotional preferences of each agent. This enables the emotional specification of a piece of music, and combines what are normally two separate processes: algorithmic composition and computer expressive performance. The iterative nature of the evolutionary method in our multi-agent system allows the generation of complex layers of expressive performance which generate sufficient levels of deviation to produce novel compositions.

1. Introduction

Evolutionary Computation is based on iterative processes in a population, for example learning, evolution and communication. Such processes are often inspired by biological and cultural development, and guided in ways to produce desired results. A common example is Genetic Algorithms which have been applied to algorithmic composition (Horner and Goldberg 1991; Jacob 1995; Miranda 2001). One recent application of Evolutionary computation has been for the examination of the mechanisms through which musical cultures might develop in simulated worlds populated by communities of artificial agents (Miranda 2002). Music's origins and evolution were examined based on the cultural norms that may develop under a collection of social or psychological constraints. An improved insight into the basic processes of musical origins and evolution can be of significant use to musicians seeking previously untapped approaches for new musical work. (Miranda et al 2003)

An early example was given in Miranda (2002 - a multi-agent system of software agents with simplified artificial singing biology. The agents interacted and a shared repertoire of tunes emerges in the population. This demonstrated how a multi-agent system without explicit musical rules could still generate a shared repertoire. As a result of Miranda's original study, Kirke and Miranda have been developing "Social Composition Systems" - applications based on multi-agent systems. In this paper, we utilize such an approach for the generation of music to develop a system we call Empath. Rather than equipping agents with singing models, agents in Empath are given an emotional state and a performance model. This model is an abstract rather than biologically-inspired, and is drawn from the field of Computer Systems for Expressive Music Performance (CSEMPs) (Kirke and Miranda 2009). Empath agents communicate their emotions, and a simple composition process emerges based on the expressive performance between agents, agents' emotional preferences, and the population's emotional profile. Apart from enabling the

emotional specification of a piece of music, this approach utilizes expressive performance rules to generate compositional elements.

2. The Agent Society

The Empath system (E.M.P.A.T.H. stands for "Expressive Music Performance through Affective Transmission Heuristics") is a society of emotional artificial agents – shown in Figure 1. Each agent in Empath has its own affective (emotional) state, labelled as: Happy, Sad, Angry or Neutral, and its own tune. Agents each start with their own "inherent" tune. Agent's randomly interact with each other, and attempt to communicate their affective state through a musical performance shaped by that affective state – i.e. a happy, sad, angry, or neutral performance. The performances of agents are generated using an implementation of the *Director Musices* Emotional Performance system (Friberg et al 2006) which will be described in more detail later. When an agent hears a tune performed by another agent, it will estimate the affective nature of the performance, and if that matches its own current affective state, the agent will add the performance to its own tune. It may also adjust its own affective state based on what it has heard, just as humans often do (Schmidt and Trainor 2003). This emotional communication mechanism is partially inspired by the idea that music was an evolutionary precursor to language in humans, and particularly adept at communicating emotions (Mithen 2005; McDermott 2008).

Agents also have an Affective Tendency – for example they may tend to be "Happy Agents" or "Sad Agents", and they have a probability of entering that affective state for no reason. Affective tendency and apparently random change in emotion mirrors our subjective experience and experimental evidence (Lorber and Slep 2006; Diener 2008). The probability of this spontaneous change is an adjustable parameter of Empath, called the Affective Tendency Probability.
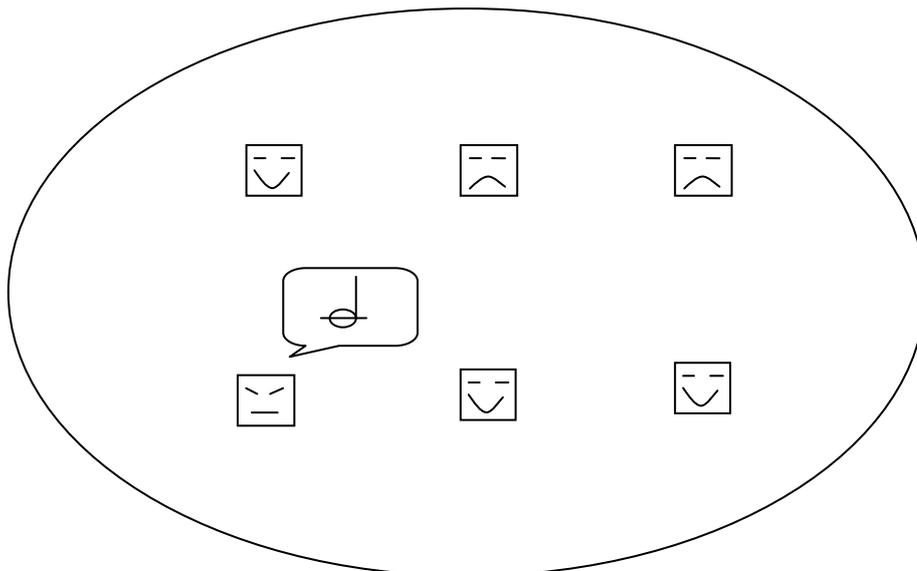

Figure 1: Society of six Empath agents – currently three happy, two sad, and one angry. The angry agent is performing its tune based on its emotional state.

Empath runs in a series of "cycles". At each cycle an agent randomly selects another agent to attempt to communicate its affective state to. In this way each agent builds a growing internal tune from the performances of other agents. The number of cycles run is defined by the user or by a stopping condition. Before starting these cycles, agents' tunes and affective states are initialised. A cycle of Empath is described below and represented in Figure 2:

1. A random agent is chosen to perform, say agent A. The agent performs its tune, adjusting it based on its affective state.
2. An agent is selected randomly, say B, to listen to A's tune
3. Agent B estimates the affective content of A's performance, and - if the affective content matches B's current affective state - adds the performance to its own tune.
4. Agent B updates its affective state either based on its effective estimate of the A's performance, or upon its own affective tendency, depending on the Affective Tendency Probability.
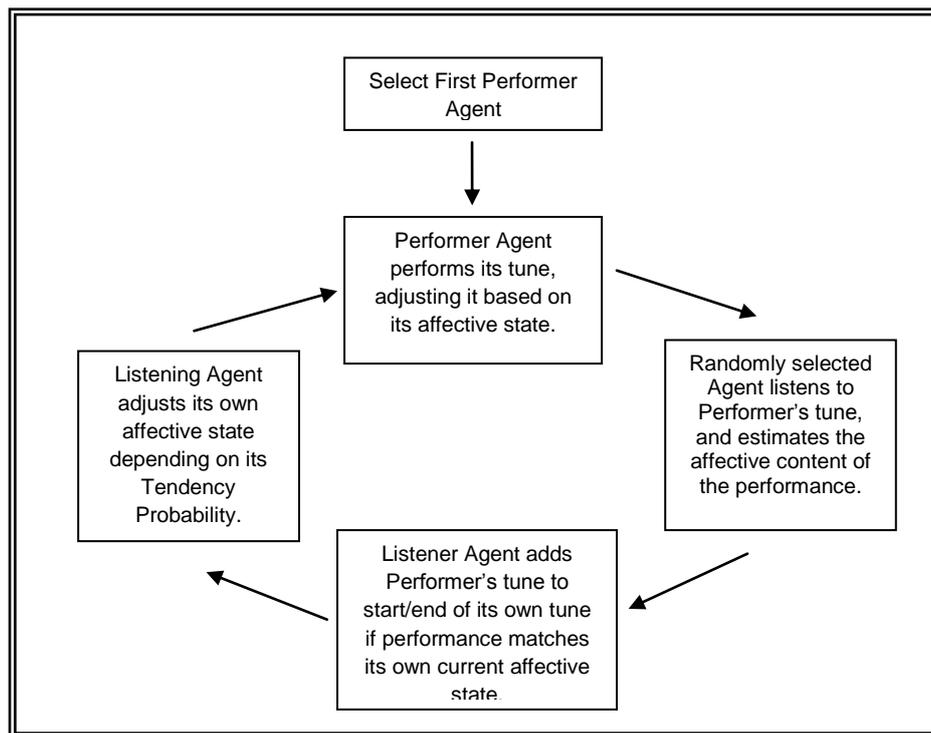5. Return to step 1

Figure 2: Cycle

3. Agents

An agent has the following information stored (summarised in Figure 3 below):

- Tune – a MIDI file representing a monophonic tune
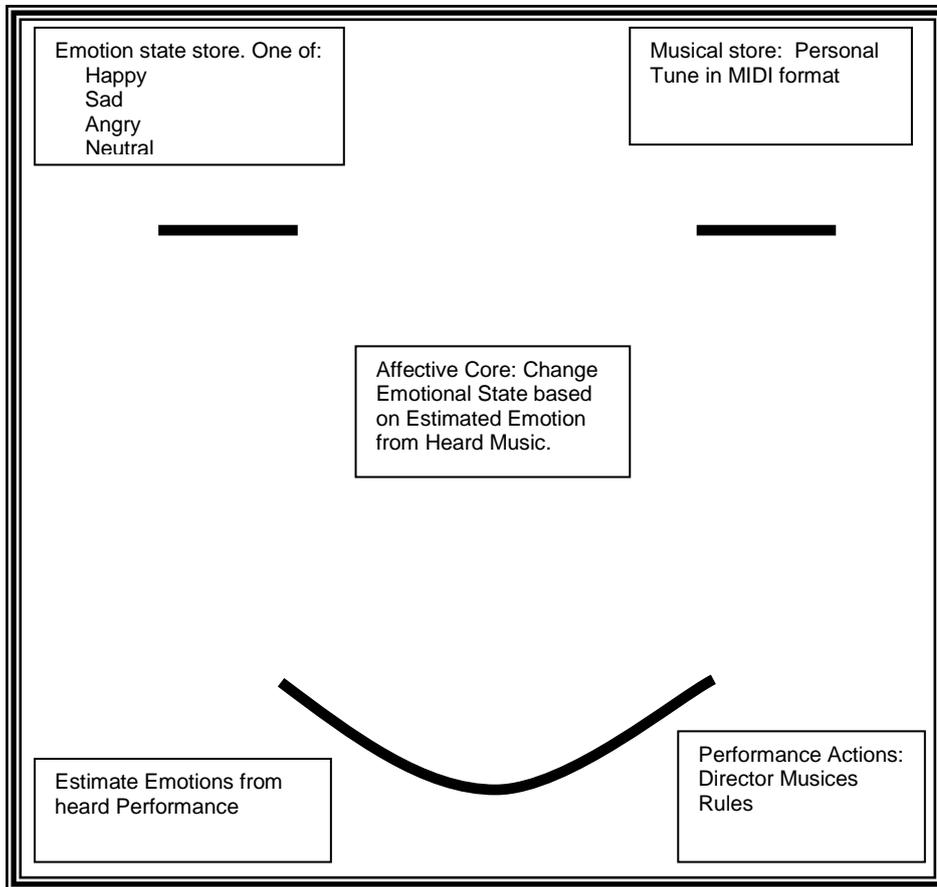- Affective State - Happy, Sad, Angry or Neutral

Figure 3: Agent Elements

An agent has the following abilities, which are also included in Figure 3:

- Perform - perform its tune, adjusting it according to its affective state.
- Listen - listen to a tune, and optionally add it to its own.
- Affective Estimation – estimate the affective content of a tune

The next few sections describe an agents' stored information and abilities.

4. Agent Tune

An agent's tune is a monophonic MIDI object. An agent can be provided with a ready-composed initial short phrase by the user or a motif generation system can be used. The selection of short phrases can be provided by the user as a single piece of

music (e.g. a MIDI file) which is then auto-chopped up by Empath – an application of the cut-up technique. In this paper we will utilize the phrase generation algorithm, rather than using cut-up, or providing pre-composed phrases.

The generation algorithm allows tunes of a randomly generated length up to a maximum length (in number of notes) set by the user. Each note in the tune has a random duration up to a value set by the user – called the base duration. The default value for this is two seconds. For simplicity, the onsets of the notes are calculated so that each note starts immediately after the previous note ends – i.e. with no rests. Included in the initial tune generation is an optional quantization process (which by default is switched on) which quantizes note durations using a user-set tick – with the default 0.0625 seconds.

The generation of pitch is through an alleatoric composition method (Roads 1996) – a uniform walk process with jumps (Xenakis 1963). By default there is a 25% probability of a jump of up to eight semitones. Otherwise at each step there is a 50% probability of the note moving up by one semitone and a 50% probability of the note moving down 1 pitch. The start point for the tune is parameterized by the Phrase Centre and Phrase Bounds – set by default to 60 (middle C) and 20 (20 semi-tones) by default. The first note will be uniformly picked within Phrase Bound semitones of the Phrase Centre, and then the walk process with jumps will be used. The amplitude generation is a uniform distribution between MIDI values 50 and 70 (D below middle C to B-flat above middle C). All these default values can be set to user values as well.

5. Agent Affective State

The agent can have one of four affective states: Happy, Sad, or Angry and Neutral. Neutral corresponds to no expressive performance. One reason for choosing the Happy, Sad and Angry states is that these three states are amongst those shown by the Director Musices Emotion System to be communicable using Expressive Performance rules (Bresin and Friberg 2000). The second reason for selecting these 3 involves the agents' ability to interpret each others' affective states through music. Work has been done by a member of the Director Musices team (Friberg 2004) on automatically detecting intended emotional states in expressive performances, using the same type of musical features generated by Director Musices. Friberg showed some success in detecting the 3 states intended by performers - Happy, Sad, and Angry.

Users can set the affective tendencies of each individual agent. As well as setting agents' affective tendency, their Tendency Probability defines the probability of an agent entering into its affective tendency state. For example: if an agent has a 60% Sad Tendency Probability, then during each cycle, that agent has a 60% chance of becoming sad, no matter what music it hears.

6. Agent Performance

When agents perform their tune, they adjust their performance based on their affective state. The adjustments made to the tune based on affective state involve one or more of note onset, amplitude and note duration. If, for example, an agent is happy, it will generate a different performance of the same tune compared to if it is sad. These adjustments are based on the Director Musices Music Emotion system and will be detailed later.

Before describing agent performance, it is helpful to discuss computer performance in more depth.

6.1 Computers and Performance

From the end of the 1980s onwards there was an increasing interest in automated and semi-automated Computer Systems for Expressive Music Performance (CSEMP). A CSEMP is a computer system able to generate expressive performances of music. For example software for music typesetting will often be used to write a piece of music, but some packages play back the music in a relatively robotic way – the addition of a CSEMP enables a more realistic playback. Or an MP3 player could include a CSEMP which would allow performances of music to be adjusted to different performance styles.

How do humans make their performances sound so different to the so-called "perfect" but robotic performance a computer would give? In this paper the strategies and changes which are most often not marked in a score and are more relatively notated, but which performers apply to the music will be referred to as expressive Performance Actions. Two of the most common performance actions are changing the Tempo and the Loudness of the piece as it is played. These should not be confused with the tempo or loudness changes marked in the score, like accelerando or mezzo-forte, but to additional tempo and loudness changes not marked in the score. For example, a common expressive performance strategy is for the performer to slow down as they approach the end of the piece [Friberg and Sundberg 1999]. Another performance action is the use of expressive articulation – when a performer chooses to play notes in a more staccato (short and pronounced) or legato (smooth) way. Another factor influencing expressive performance actions is Performance Context. Performers may wish to express a certain mood or emotion (e.g. sadness, happiness) through a piece of music. Performers have been shown to change the tempo and dynamics of a piece when asked to express an emotion as they play it (Gabrielsson and Juslin 1996). For a discussion of other factors involved in human expressive performance, we refer the reader to (Juslin 2003).

6.2 Director Musices

The approach we utilize here is based on Director Musices (Sundberg et al. 1983)(Friberg et al. 2006), which has been an ongoing project since 1982. Researchers including violinist Lars Fryden developed and tested performance rules using an analysis-by-synthesis method (later using analysis-by-measurement and studying actual performances). Currently there are around 30 rules which are written as relatively simple equations that take as input music features such as height of the current note pitch, the pitch of the current note relative to the key of the piece, or whether the current note is the first or last note of the phrase. The output of the equations defines the Performance Actions. For example the higher the pitch the louder the note is played, or during an upward run of notes, play the piece faster.

Another Director Musices rule is the Phrase Arch which defines a "rainbow" shape of tempo and dynamics over a phrase. The performance speeds up and gets louder towards the centre of a phrase and then tails off again in tempo and dynamics towards the end of the phrase. The Punctuation rule adds micro-pauses at the end of phrases. For many Director Musices rules some manual score analysis is required – for example harmonic analysis and marking up of phrase start and ends. Each equation has a numeric "k-value" - the higher the k-value the more effect the rule will have and a k-value of zero switches the rule off. The results of the equations are added together linearly to get the final performance.

As well as humanization of the performance, Director Musices is also able to implement emotional expression (Bresin and Friberg 2000), drawing on work from Gabrielsson and Juslin (1996). Listening experiments were used to define the k-value settings on the Director Musices rules for expressing emotions. The music used was a Swedish nursery rhyme and a computer-generated piece in minor mode written using the Cope (2005) algorithmic composition system in the musical style of Chopin. Six rules were used from Director Musices to generate multiple performances of each piece. Subjects were asked to identify a performance emotion from the list: fear, anger, happiness, sadness, solemnity, tenderness or no-expression. As a result parameters were found for each of the six rules which mould the emotional expression of a piece. The parameters for "tenderness" would be: inter-onset interval is lengthened by 30%, sound level reduced by 6dB, and two other rules are used: the Final Ritardando rule (slowing down at the end of a piece) and the Duration Contrast rule (if two adjacent notes have contrasting durations, increase this contrast).

6.3 Agent Performance Abilities
An Empath agent performance is based on the following Director Musices rules:

1. *Tone IOI* (Inter-Onset Interval – the time between first note start and second note start) – Increases or decreases the tempo.
2. *Sound Level* - Increases or decreases the loudness.
3. *High Loud* – Increase loudness as pitch increases
4. *Punctuation* – add a micro-pause and duration change at phrase boundaries to emphasise phrasing.
5. *Duration Contrast* – relatively short notes are shortened and relatively long notes are lengthened.
6. *Duration Contrast Articulation* – inserts a micro-pause between adjacent notes if the first note has an IOI between 30 and 600 milliseconds.
7. *Phrase Arch* – adds an accelerando to the start of a tune and a ritardando to the end. Similarly adds a crescendo to the start and a decrescendo to the end.

These are all the rules identified in Bresin and Friberg (2000) as enabling emotional expression through Director Musices, except one. The one rule excluded from this list is final ritardando. This rule relates to the final slowing down that is often heard when a performer reaches the end of a piece of music. This rule is not utilised here because Director Musices rules in our system are applied repeatedly as tunes build up in an agents' memory; whereas the final ritardando rule should only be applied once and at the end of the composition. Applying it repeatedly every time an agent

performs will lead to multiple final ritardandos appearing in the middle of the piece. All of the other rules (except Phrase Arch) are based on the note level details of the piece rather than its larger scale phrase structure, and can be viewed as emphasising their effects during their re-application as a tune iteratively expands in Empath.

The only other structural rule in emotional expression – Phrase Arch – is actually applied repeatedly at multiple levels in Director Musices. So it is more reasonable to use such a structural rule in the repeated interactions of agents. In the case of all the Director Musices rules utilized, a natural question is whether such rules retain their purpose and meaning when repeatedly iterated as is done in Empath. This will be addressed in the experiments in the next chapter, but is also touched upon later in this section. Bresin and Friberg (2000) examined what selection of rules, and what parameter values for the rules, were appropriate to express certain emotions through performance. The results are shown in Table 1.

	Sad	Happy	Angry	Neutral
Tone IOI	Lengthen 30%	Shorten 20%	Shorten 15%	N/A
Sound Level	Decrease 6dB	Increase 3dB	Increase 8dB	N/A
High Loud (increases loudness for higher pitches)	N/A	k = 1.5	N/A	N/A
Punctuation	N/A	k = 2	k = 2	N/A
Duration Contrast	k = -2	k = 2	k = 2	N/A
Duration Contrast Articulation	N/A	k = 2.5	k = 1	N/A
Phrase Arch	Level 5: k = 1.5 Turn = 0.25 Level 6: k = 1.5 Turn = 2	N/A	Level 5: k = -0.75 Turn = 0.5 Level 6: k = -0.75 Turn = 0.25	N/A

Table 1: List of Director Musice Rule Parameters to Express Emotional States

There are a few observations to be made from this table. First of all – not all rules are used for all of the emotional expressions. Secondly, sound level increases are defined in decibels. This is a relative measurement, and loudness is not defined in MIDI as a decibel measurement. Different MIDI devices will give different decibel changes for MIDI loudness changes. To deal with this issue we benchmarked against a response averaged over two MIDI instruments: a Yamaha Disklavier and a Bosendorfer SE290. The reason for this choice is threefold: (1) this version of Empath is only tested and utilized on piano; (2) these two instruments are commonly used MIDI pianos; and (3) the MIDI loudness value vs. decibel readings for these instruments are calculated in Goebl and Bresin (2003). These values are used in the current version of the Empath.

A piece of music has a number of levels of meaning – a hierarchy. Notes make up motifs, motifs make up phrases, phrases make up sections, sections make up a piece. Each element - note, motif, etc - plays a role in other higher elements. Human

performers have been shown to express this hierarchical structure in their performances. Phrase Arch is defined at two hierarchical levels in the table: Levels 5 and 6. Examples of these grouping levels are shown in Figure 3. In the original Director Musices Emotion System, each grouping on Level 6 will have its own phrase arch generated for it using the Level 6 parameters in Table 1, and each grouping on Level 5 will have its own phrase arch generated for it using the Level 5 parameters in Table 1. However in our system, the application of Phrase Arch happens in a different way, as the rule is applied during and as part of composition. In (Kirke and Miranda 2008) a similar multi-agent methodology was presented which allowed a Phrase Arch-type rule to be applied to multiple levels in the hierarchy without the need for the type of structural analysis shown in Figure 3. Such an approach generates a structurally meaningful set of accelerandi and ritardandi.

Figure 4: Examples of Phrase Arch Levels 5 and 6 in Schumann's Träumerei (Friberg 1995)

To understand how such an approach is implemented, consider a Multi-agent System (MAS) of two agents A and B. Suppose their motifs would both be classed as Level 7 motifs in the type of analysis shown in Figure 3 (these would be short motifs of three or five notes in length). If A performs its tune, then B will add that to the start of the end of its tune. B's tune will then be six to ten notes long, more of a Level 6 length tune. When B performs its tune, A will add the tune to the start of its own, thus A's tune becomes nine to fifteen notes long, closer to a Level 5 tune. This shows how the process of agent communication and tune addition causes a hierarchical building up of tunes through different levels. In addition to this length extension process, at each interaction a Phrase Arch is added by the performing agent over the tune it performs. Thus Phrase Arch is being applied hierarchically to the composition / performance as it is being generated in each agent. However, the Director Musices Emotion System only defines its parameters for levels 5 and 6, and the composition method of the MAS will extend beyond such levels after only a few iterations. So in order to allow the application of the Phrase Arch rule to affective generation, a compromise was made whereby a single Phrase Arch rule was applied each time an agent performs; and the parameters of that rule are the average of the parameters for Levels 5 and 6 in Table 1. For "Sad" this gives k = 1.5 and Turn = 0.25; for "Angry" k = -0.75 and Turn = 0.375. Such an averaging and application will be further discussed in the results section.

7. Agent Listening

When an agent listens to another agent's tune, it analyses four global features of the tune - loudness, tempo, level of legato / staccato, and tempo variability - and attempts to detect the emotion in the tune's performance. These features are based on the features from the emotion detection work in Friberg (2004). However, where Friberg uses a fuzzy logic system to detect emotions in music, Empath utilizes a simple linear regression model based on the same features as Friberg, to extract affective states from the musical features. If the performance is estimated as having the same affective content as the listening agent's current state, the performed tune is added to the listening agent's tune. The tune will be added either to the end or beginning of the listening agent's own tune – as defined by a parameter set by the user (e.g. 50% probability either way). Note that agents store *performances* – thus even the smallest micro changes added by a performing agent will become part of the performed tune, and be stored by the listening agent.

Friberg (2004) proposes the features in Table 2 as cues for detecting the emotionality of a tune's performance. This table can be compared to Table 1 which describes the Director Musices rule parameters for generating emotional cues. One element of this Table is not used in the current version of our system. This is the "sharp timbre" element, since this MAS focuses on Piano music in its experiments.

Emotion	Musical Cues
Anger	Loud Fast Staccato Sharp timbre
Sadness	Soft Slow Legato
Happiness	Loud Fast Staccato Small tempo variability

Table 2: List of Music Performance Cues from (Friberg 2004)

These cues are defined by Friberg more precisely based on a number of features: (1) sound level, (2) tempo, (3) articulation, (4) attack rate and (5) high-frequency content. Of these five features only the first three are used in Empath, due to the fact that we focus on MIDI Piano. When an agent hears a performance, it calculates the sound level by averaging the MIDI Loudness parameter across the tune; the tempo is calculated by taking the reciprocal of the average inter-onset interval; the articulation is calculated using the Relative Pause Duration (Schoonderwaldt et al 2001) which is defined as the distance between the next tone, and the end of the previous tone, divided by the inter-onset interval between the two tones. This is designed to measure how legato or staccato a tone is. These features are summarized in equations (1), (2) and (3).

$$Sound\ level = mean(loudness(Tune)) \tag{1}$$
$$Tempo = 1\ /\ (mean(IOI(Tune))) \tag{2}$$
$$Articulation = mean(Pause\ Duration\ /\ IOI) \tag{3}$$

Looking at Table 2 it can be seen that there is one more feature not covered by these equations: tempo variability. This is estimated in Empath by calculating the standard deviation of the IOI in equation (4).

$$Tempo\ Variability = std(IOI(Tune)) \tag{4}$$

In (Friberg 2004), features are combined with visual cues in a Fuzzy Logic system to detect emotional state. In our approach a simpler model is used – linear regression. This model is shown in equation (5).

$$Affective\ State = a * Sound\ level + b * Tempo + c * Articulation +$$
$$d * Tempo\ Variability + e \tag{5}$$

The calculation of the parameters a, b, c and d was performed as follows. 2048 random MIDI files were generated of up to 8 beats. They were generated using the same initial tune algorithm that agents use. For each of the 2048 tunes, the Rules from Table 1 were applied (with the already discussed adjustments specific to our system) to generate each emotional state: Happy, Sad, Angry and Neutral – thus giving 2048 MIDI files for each emotion = 8192 MIDI files. Then the features (1), (2), (3) and (4) were calculated for each of the 8192 files. The 3 affective states Happy, Sad and Angry were each given a numeric index: 3, 2 and 1. This allowed for a linear regression to be set up between features (1), (2), (3) and (4) (as applied to the 8192 files) and the affective state index. This index was known, because it was known which of the 8192 files had which of the Director Musices rules applied to it for emotional expression. The resulting model from the linear regression is shown in equation (6).

$$Affective\ State = -0.04 * Sound\ level + 2.8 * Tempo + 0.14 * Articulation -$$
$$0.1 * Tempo\ Variability + 4.1 \tag{6}$$

In applying the model, the actual effective state is calculated using (6) by rounding the result to an integer and bounding it between 1 and 4. So if the calculated effective state is greater than 4, then it is set to 4. If the calculated effective state is less than 1 then it was set to 1. When this was tested against a further five random sets of 8192 tunes the average error in affective state was 1.3. This translates to an accuracy of 68%. On first glance this would seem to be too low to be utilizable. However a key point needs to be taken into account. Human emotion recognition levels in music are not particularly high – for example tests using a flute didn't surpass 77% accuracy (Camurri et al. 2000). Although these accuracies are not precisely comparable (since they involve different instruments and emotion sets) they give a helpful reference point. For example - implementing a recognition system with perfect accuracy would lead to a non-social model – i.e. an agent model which was far separated from human musical groups. Furthermore using a benchmark of the order of 77%, we consider 68% to be a high enough accuracy with which to implement the model. A useful perspective is obtained by considering it to be 88% (= 68/77) as accurate as humans. We will call this model the "Empath Linear Model". It is interesting to note that the model (6) is most influenced by Tempo, and is only

minimally influenced by Sound level. This may help to explain some of the difference between the Linear Model and human recognition rates.

8. Empath Experiments

We will now run a series of experiments to demonstrate properties of Empath.

8.1 Effects of Emotional Tendency

The first demonstration will be to show how the emotional tendency of the agent population can be used to change the emotional profile of the resulting composition population. Experiments were run with 16 agents. They were left communicating until an agent existed with a piece of a minimum length (1800 beats), or after a maximum number of cycles were completed (1536 cycles). 10 Experiments were run in each of three conditions: all agents having Angry tendency, all having Sad tendency or all having Happy tendency. The affective tendency probability was set to 50% (i.e. 50% chance of an agent switching to its affective tendency spontaneously). The average emotional value (1 = Angry, 2 = Sad, 3 = Happy, 4 = Neutral) of all agents' *tunes* (not of the agent's affective state) at the end of cycles was averaged across the 10 experiments and is shown in Table 3 to 1 decimal place. This was estimated using the Linear Model. The values in parentheses are the Coefficients of Variation – which is the Standard Deviation divided by the mean – a measure of spread of the values. The values in brackets below this are the percentage of the population who are angry, happy, sad and neutral at the end of the cycles (neutral is left out if there are 0% neutral agents).

100% Angry (Value 1)	100% Sad (Value 2)	100% Happy (Value 3)
1.1 (21%)	2.1 (11%)	3 (3%)
[97%, 0%, 3%]	[0%, 89%, 11%]	[0%, 0%, 96%, 4%]

Table 3: Effects of 100% Affective Tendencies

It can be seen in Table 3 that the average performance emotion value of the population of tunes at the end of the evolution corresponds closely to the emotion tendency of the agents. This is as would be expected, but was by no means guaranteed, due to the recursive nature of the performance applications as agents share tunes. It is interesting to note that different emotional tendencies seem to have different levels of "stability". Angry agents generate final composition / performances which the Linear Model estimates as being more emotionally diverse. Whereas this happens less so for sad populations and happy populations produce hardly any emotional diversity at all. This is further investigated in later experiments. It is interesting to note, that the stability is loosely correlated with the Valence (positivity) of the emotion. Emotionally more positive populations seem to be more stable, though this may be just a quirk of the Empath Linear Model.

This use of tendencies is further demonstrated in Table 4. In this Table two-thirds of all agents were given the same emotional tendency, and the other third of the population were given equal amounts of the remaining tendencies. Once again, ten

runs were done of each up to 1800 beats or 1536 cycles. The resulting average emotion value calculated by the linear model is given, together with its coefficient of variation in parentheses; and the number of agents in each state [angry, sad, happy, neutral] is listed.

Angry 67%, Sad / Happy 16.5%	Sad 67%, Angry / Happy 16.5%	Happy 67%, Angry / Sad 16.5%
1.5 (54%)	2.3 (22%)	2.9 (9%)
[72%, 11%, 17%, 1%]	[3%, 63%, 35%]	[1%, 8%, 89%, 3%]

Table 4: Effects of reducing Affective Tendency Majority

(Note that in Table 4 and the following tables, percentages have been rounded to integers for clarity.) It can be seen that the 67% angry population generates on average more angry performances than the 67% sad or 67% happy populations. However, if the average tune emotion value for the 67% angry population (= 1.5) in Table 4 is compared to the value for 100% angry population (= 1.1) in Table 3, it can be seen that the average emotionality has been pulled more towards other emotions, due to the influence of the sad and happy 16.5% subpopulations. The variation for the 67% angry (54%) has doubled compared to 100% angry. The effect of subpopulations can be observed in a similar way in Table 4 for the 67% sad population and the 67% happy population. Both are pulled away from uniform emotionality and have at least a doubled variation.

This dilution process reaches its maximum extent if all emotional tendencies are distributed equally across the population – i.e. 33% angry, 33% happy, and 33% sad. Ten experiments were run like this. The resulting average emotionality of tunes across the population is shown in Table 5.

Angry/Sad/Happy 33%
2.3 (31%)
[17%, 38%, 45%]

Table 5: Effects of Equal Distribution of Affective Tendencies

The result demonstrates a significant increase in variation over Table 4 for the sad and happy majority columns, and a significant decrease for the angry majority column. The increase in variation is as would be expected for happiness and sadness. The decrease relative to angry in Table 4 is probably due to the stabilising effect on the population of increasing the amount of happy and sad tendency agents (these have been shown to be more stable emotions in iterations). These results demonstrate two things – first that this 33% experimental state is significantly different to both the 100% and 67% populations – giving somewhat the expected and hoped for behaviour. Secondly, it further demonstrates the definite bias in Empath towards the happy state.

This "optimistic" bias in Empath explains why the stability increases for populations with happy majorities. It is not clear if this is a result of the Linear Model or

something to do with the performance / composition process. Possible reasons are that the linear model is not very interested in Tempo Variability (see Equation (6)), an element which Happiness does not utilize as significantly as the other emotions. It does not utilize Tempo Variability as significantly because it does not use Phrase Arch. Another way to conceptually visualize this dominance situation is that agents have a different emotional confusion tendency than humans do. Humans find certain emotions in music hard to distinguish – though angry and happy, or sad and happy, are not emotions they would usually confuse. However the Empath agents have a different confusion profile – having a higher tendency to confuse angry with happy / sad than humans would.)

8.2 Other Population Elements

Another set of variations examined was the population size. Experiments were run with population sizes of 8, 32 and 64 (in addition to those run for the 16 agent populations). These were done for the 33% angry, happy and sad tendency, each up to 1800 beats or 1536 cycles, and a tendency probability of 0.5. The results are shown in Table 6.

8 Agents	16 Agents	32 Agents	64 Agents
2.7 (14%)	2.3 (31%)	2.2 (36%)	2.1 (37%)
[4%, 24%, 73%]	[17%, 38%, 45%]	[24%, 37%, 39%]	[24%, 36%, 39%]

Table 6: Effects of Changing Population Size

The only pattern which can be discerned here is that the larger the agent group, the less dominance happy tunes have over sad and angry tunes. Experiments were also done with changing the affective tendency probability. This was done with 16 Agents with 33% of Angry / Happy / Sad tendency. The results are shown in Table 7.

AT 0	AT 0.25	AT 0.5	AT 0.75	AT 1
2.3 (28%)	2.3 (31%)	2.2 (35%)	2.3 (31%)	2.3 (33%)
[14%, 40%, 46%]	[17%, 38%, 45]	[23%, 38%, 39%]	[16%, 43%, 41%]	[19%, 35%,46%]

Table 7: Effects of Changing Affective Tendency Probability

This table demonstrates why an affective tendency probability of 0.5 was chosen for all of the other experiments. For this order of cycles and number of agents a probability of 0.5 (i.e. a 50% chance of the agent spontaneously switching to its emotional tendency) reduces the excessive "optimism" – i.e. the dominance of happy tunes - of Empath.

8.3 Effects of Structural Expression

One other element to be demonstrated is the structural expression results for Empath. One of the key structural performance elements in human performance is slowing between boundaries in a piece. We will examine Agent 6's tune from run ten in the 100% angry tendency experiment at the start of this section. This tune is chosen because it is repetitive, and more clearly shows deviations due to performance rather than initial tune generation. We can look at how the tune builds up from an interaction history from the run for Agent 6, as shown in Table 8. The table shows the cycles during which agent 6 was performed to, the agent that performed to it in that cycle, the length of agent 6's tune (number of MIDI notes) at the end of that cycle, and the affective state agent 6 was in at the end of that cycle as a result of hearing the performed tune. During some interaction rows, agent 6's tune does not increase in size because the tune performed to it did not match its affective state at the start of the cycle. It can be seen that the agent's tune is built up from combining six tunes with Agent 6's initial tune.

Cycle	Agent Listened to	Tune Length	Affective State
1	-	2	2
2	5	7	2
11	8	7	3
24	1	12	2
31	1	17	2
38	9	17	3
39	3	17	1
40	11	17	3
60	5	17	1
67	14	17	4
81	13	17	3
86	12	17	3
87	8	22	3
98	9	22	3
104	5	22	1
133	7	99	2
156	15	232	2
160	4	232	1
166	2	232	1

Table 8: Agent 6 Interactions and States

Figure 5 shows a piano roll of Agent 6's final tune, together with a smoothed plot of average duration (ten notes to one data point). Firstly, it can be seen that there are 8 significant slowing points in the graph (including the start and end boundaries). This corresponds with the boundaries of the seven tunes used (Agent 6's initial tune plus the six added tunes during interaction). It can also be seen that although the tunes are repetitive, the right hand of the of the duration graph is significantly lower that the left hand. This is due to Agent 6's final interaction (with Agent 15). At this interaction Agent 6's tune goes from 99 notes to 232 notes. So this slowing expresses the boundary between Agent 6's 99 note tune and Agent 15's 133 note tune.
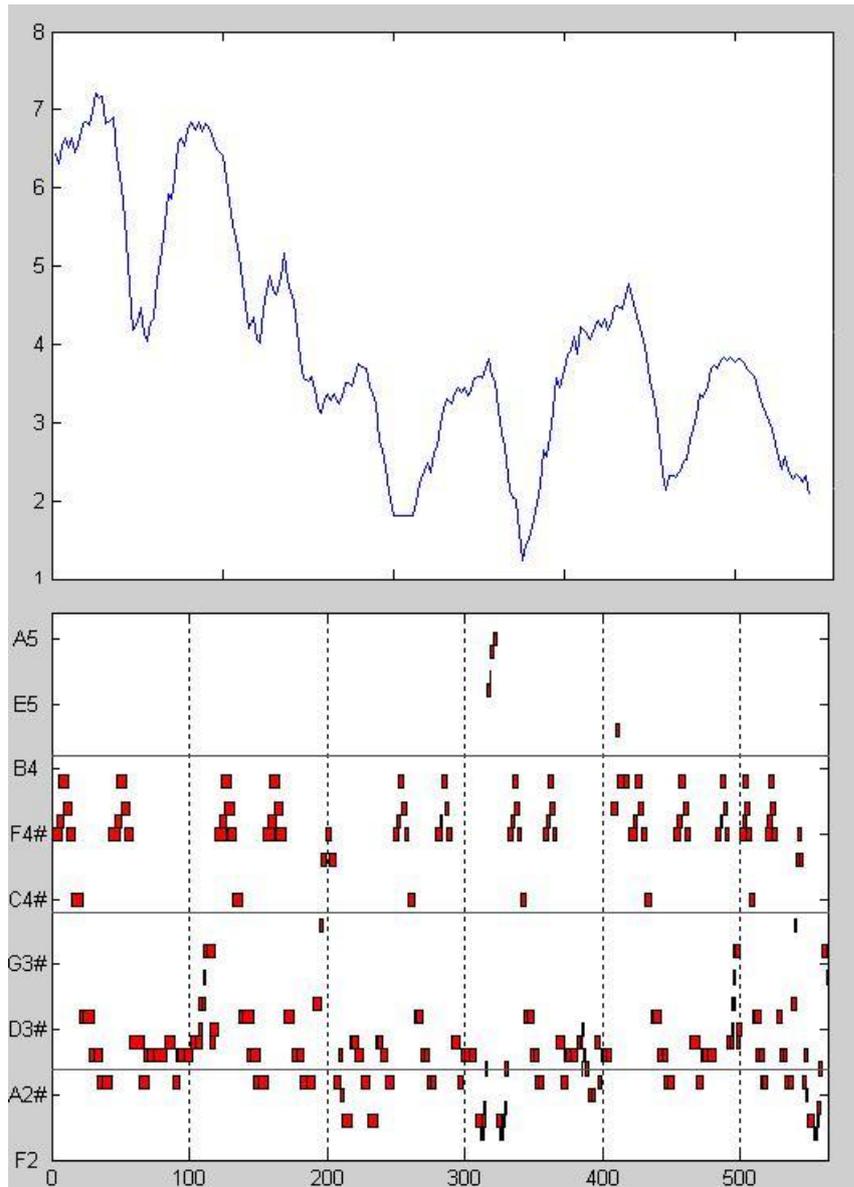
Figure 5: Structural Expression in Empath

As a final example of the use of Empath we show common music notation excerpts from Agent 1, selected from run 10 of the 100% Angry / Sad / Happy tendency of the first experiments in this section. The excerpts are shown in order of increasing tempo in Figures 6, 7 and 8.

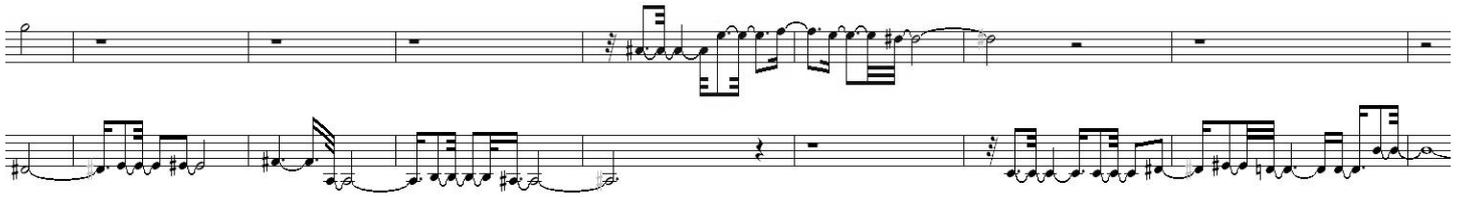
Figure 6: Agent 1 – Sad 100% - Music Excerpt

Figure 7: Agent 1 – Angry 100% - Music Excerpt

Figure 8: Agent 1 – Happy 100% - Music Excerpt

It can be seen that the tempi of the three excerpts are in line with Table 1's tempo changes based on emotion, as expected.

9. Conclusions

We have introduced a multi-agent composition and performance system called Empath (Expressive Music Performances through Affective Transmission Heuristics). It is an artificial social composition system based on evolutionary musicology models, and generates a population of performance / compositions. Rather than equipping agents with singing models, agents in Empath are given an emotional state and a performance model. This model is an abstract rather than physically based, and is drawn from the field of Computer Systems for Expressive Performance. We have shown how Empath agents are able to attempt to communicate their emotions, and how based on this a simple performance / composition process emerges. We have also shown that the final population of compositions has a distribution which can be significantly affected by the setting of the initial population affective tendencies and other Empath parameters. Apart from enabling the emotional specification of a piece of music, this approach also leads to the generation of a piece of music which expresses its motivic and larger scale structure.

Concerning/Regarding future work, one thing that would be useful for Empath would be formal listening tests. The Linear Model used to evaluate experiments is sufficiently accurate to give strong indicative results, but has limited accuracy beyond that. A related element is an investigation of alternatives to the Linear Model. Other models are possible – including polynomial, neural network, or fuzzy logic models. A key element which could be extended is the initial tune generation in Empath. The alleatoric method was useful as a proof of concept for social emotional performance/composition. However, the initial agent motifs could be generated using any number of different compositional systems – not just alleatoric - and they do not

need to be independent. Also, emotional composition methods (e.g. pitch and key changes) could be added to the emotional performance expression between agents, implemented as transformations or communication errors during agent interaction. For example, a Happy agent may transpose its tune into a minor key; or an Angry agent may perform its tune fast and with less accuracy. Furthermore the methodology used to choose which agent to interact with could be extended to give the transformations greater musical intelligence. This musical intelligence could be further extended using social networks – where agents have a variable "opinion" of each other (Kirke 1997) – and use this to motivate or restrict interactions in certain ways. Such social networks may also be able to generate agent social hierarchies which can be used to generate more meaningful musical hierarchies in the final compositional population.

Despite the large potential for future work, Empath is a clear demonstration of social composition systems, and has been shown to implement one approach – an affectively constrained methodology – and to generate tunes which are expressively performed.

References

Bresin, R. and Friberg, A. 2000. "Emotional Coloring of Computer-Controlled Music Performances." *Computer Music Journal*,24(4): 44-63.

Camurri, A., Dillon, R., and Saron, A. 2000. "An Experiment on Analysis and Synthesis of Musical Expressivity." *Proceedings of 13th Colloquium on Musical Informatics, L'Aquila, Italy, September 2000.*

Cope, D. 2005. *Computer Models of Musical Creativity*. Cambridge, MA, USA.

Diener, E. 2008. "Myths in the science of happiness, and directions for future research." In *The science of subjective well-being*. New York: Guilford Press, pp. 493-514.

Friberg, A. 1995. "Matching the Rule Parameters of Phrase Arch to Performances of`Träumerei': A Preliminary Study." *Proceedings of the KTH Symposium on Grammars for Music Performance, Stockholm, May 1995*. Stockholm: Royal Institute of Technology, KTH, pp. 37-44.

Friberg, A. 2004. "A fuzzy analyzer of emotional expression in music performance and body motion." *Proceedings of 2004 Music and Music Science*.

Friberg, A., Bresin, R., and Sundberg, J. 2006. "Overview of the KTH rule system for musical performance." *Advances in Cognitive Psychology*,2(2): 145-161.

Gabrielsson, A. and Juslin, P. 1996. "Emotional expression in music performance: between the performer's intention and the listener's experience." *Psychology of Music*,2468-91.

Gabrielsson, A. 2003. "Music performance research at the millenium." *Psychology of Music*,31221-272.

Gabrielsson, A. 2003. "Music performance research at the millenium." *Psychology of Music*,31221-272.

Goebl, W., Bresin, R., 2003, "Measurement and Reproduction Accuracy of Computer Controlled Grand Pianos." *Journal of the Acoustical Society of America,* 114(4), 2273–2283.

Horner, A. and Goldberg, D. 1991. "Genetic Algorithms and Computer-Assisted Music Composition." *Proceedings of 1991 International Computer Music Conference*. USA: ICMA, pp.

Jacob, B. 1995. "Composing with Genetic Algorithms." *Proceedings of the 1995 International Computer Music Conference*.: ICMA, pp.

Kirke, A. 1997. "Learning and Co-operation in Mobile Multi-robot Systems" PhD Thesis, University of Plymouth.

Kirke, A. and Miranda, E. R. 2008. "An Instance Based Model for Generating Expressive Performance During Composition", *Proceedings of International Computer Music Conference - (ICMC2008)*, Belfast (UK).

Kirke, A. and Miranda, E.R. 2009. "A Survey of Computer Systems for Expressive Music Performance." *ACM Surveys* (in Press)

Lorber, M. and Slep, A. 2005. "Mothers' emotion dynamics and their relations with harsh and lax discipline: microsocial time series analyses." *Journal of clinical child and adolescent psychology*,34(3): 559-568.

McDermott, J. 2008. "The Evolution of Music." *Nature*(453): 287-288.

Miranda, E., Kirby, S., and Todd, P. 2003. "On Computational Models of the Evolution of Music: From the Origins of Musical Taste to the Emergence of Grammars." *Contemporary Music Review*,22(3): 91-111.

Miranda, E.R. 2001. *Composing Music With Computers*. Oxford, UK:Focal Press.

Miranda, E.R. 2002. "Emergent Sound Repertoires in Virtual Societies." *Computer Music Journal*,26(2): 77-90.

Mithen, S. 2005. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London:Weidenfeld & Nicolson.

Noble, J. and Franks, D.W. 2004. "Social learning in a multi-agent system." *Computing and Informatics*,22(6): 561-574.

Roads, C. 1996. *The Computer Music Tutorial*. Cambridge, Massachusetts, USA:MIT Press.

Schmidt, L. and Tainor, J. 2001. "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions." *Cognition and Emotion*,15(4): 487-500.

Schoonderwaldt, E. and Friberg, A. 2001. "Towards a rule-based model for violin vibrato." *Proceedings of 2001 MOSART Workshop on Current Research Directions in Computer Music, Barcelona, Spain, November 2001*. Barcelona, Spain: Audiovisual Institute, Pompeu Fabra University, pp. 61-64.

Sundberg, J., Askenfelt, A., and Frydén, L. 1983. "Musical performance. A synthesis-by-rule approach." *Computer Music Journal*,737-43.

Xenakis, I. 7-3-1963. *Formalized Music: Thought and Mathematics in Composition*. Hillsdale, NY:Pendragon Press.