A Computer Model for the Generation of Monophonic Musical Melodies

Juan Alvarado López[1] and Rafael Pérez y Pérez[2]

[1] Posgrado en Ciencias e Ingeniería de la Computación
Universidad Nacional Autónoma de México
jalvarado@uxmcc2.iimas.unam.mx
[2] Departamento de Tecnologías de la Información
Universidad Autónoma Metropolitana, Unidad Cuajimalpa
México D.F.
rpyp@rafaelperezyperez.com

Abstract. This paper describes a system that generates monophonic melodies called ERMEG (Engagement and Reflection MElodies Generator). It is based on the engagement-reflection computer model of creativity. ERMEG performs two main processes: the creation of knowledge structures and the generation of a new melody (E-R cycle). During engagement the system produces sequences of musical phrases driven by content and musical-theory constraints and avoids the use of explicit goals or predefined melody-structures. During reflection the system breaks impasses, verifies the coherence of the melody in progress and evaluates the novelty of the material produced so far and, as a result of this evaluation, it generates a set of guidelines that work as constraints during engagement.

Key words: Creativity, Music, Automatic Composition, Engagement, Reflection.

1 Introduction

This paper describes a computer program named ERMEG (Engagement and Reflection MElodies Generator) which develops monophonic melodies. It is based on the Engagement-Reflection computer model of creativity (E-R model) which has been employed to develop a computer program for plot generation known as MEXICA [1]. The main characteristics of the E-R model are:

- The user provides a set of examples that are employed to build the system's knowledge structures.
- The model includes two core processes: a generation process (known as engagement) and an evaluation processes (known as reflection). During engagement the system generates material guided by content constraints; during reflection the system evaluates the material generated so far -modifies it if it is necessary- and as a result of the evaluation adjusts the constraints that drive the generation of material during engagement. The system's outputs are the result of the interaction between engagement and reflection.

– An important characteristic of the model is that, following Gelernter [2], emotions are the glue that joins ideas during engagement. Thus, in MEXICA, emotional links and tensions between characters drive the generation of material during the unraveling of the plot [3].
– The model provides mechanisms to control the behavior of the different processes that conform the E-R cycle.

The main purpose of this work is to explore if the ideas that inform the E-R model employed for building MEXICA can be applied for musical composition of monophonic melodies. We believe that the study of the commonalities and differences between plot generation and melodies generation employing the E-R model might result in interesting information about computational creativity. In this paper we focus in describing how we implemented the E-R model for producing monophonic melodies. Our work considers that:

– A melody can be divided in sequences of notes; we refer to them as composition-phrases.
– Composition-phrases can be associated with representations of emotions and tensions.
– These representations of tensions and emotions can be employed during engagement as cue to probe memory in order to generate a sequence of musical phrases that conform a monophonic melody.
– Following the original E-R model, the user provides a set of examples to build its knowledge structures.

In this document we describe how the system builds its knowledge structures, how we associate emotion and tensions to musical phrases and how the engagement-reflection cycle works. We show some examples of melodies produced by the system and provide some discussion about the system.

2 Creation of Knowledge Structures.

ERMEG builds its knowledge structures from a set of melodies provided by the user of the system called Previous Melodies. Each Previous Melody is divided by the system into groups of notes that we refer to as composition-phrase (CP). The length of each CP is defined by the user of the system in terms of beats. That is, the user might ask the system to create one CP every four beats. So, the notes included in the first four beats of the first Previous Melody constitutes the CP1; the notes included in the second four beats of the first Previous Melody constitutes the CP2; and so on. The process is repeated for each Previous Melody. As we will explain later, Composition-Phrases are an essential element during engagement. Because the user has the possibility of modifying CP's length, it is possible to compare ERMEG's outputs for different lengths.

Each CP has a set of three characteristics we are interested about:

– Speed: Every phrase has associated a Speed which can have any of the following values: Fast, Medium and Slow. These values depend on the number of notes in the phrase and its length (in beats). For example, if we have eight notes in two beats the phrase is classified as Fast; on the other hand, if we have three notes in eight beats, the phrase is classified as Slow. The system employs a predefined table to determine these relations.
– Interval-Difference is defined as the difference (in intervals) between the first and last notes in a phrase. It can have the following values: Big, Medium, Small.
– Interval-Variation: The Interval-Variation is defined as the average distance (in intervals) between all the notes within a phrase. Its possible values are Big, Medium and Small.

We employ these three characteristics to associate to each CP a value of tension and emotion.

2.1 Tension and Emotion

In the literature one finds different studies about music and emotion. For example, Dalla Bella [4] points out how the tempo strongly influences the type of emotions triggered by a melody: fast tempos tend to evoke happiness and slow tempos tend to evoke sadness. Khalfa et al. [5] have performed experiments that illustrate that neither the tempo nor the rhythm alone generate strong emotions; it is the combination of both, rhythm and melody, that trigger emotions in people. Juslin and Sloboda [6] have shown that tonal music in high pitches and fast tempo can generate happiness; that big variations (leaps) of pitch and high pitches can generate excitation or anxiety; that low pitches and slow tempo can generate sadness.

Based on these works, we decided to employ Speed, Interval-Difference and Interval-Variation to associate emotional and tensional characteristics to our musical phrases. The current version of the system employs one type of emotion -which represents a continuum between happiness and sadness- and one type of tension -which represents a continuum between anxiety and calmness. (We are aware that emotions in music are more complex than this. But for our first prototype we have decided to use a simple representation of emotions). The system calculates the tension and emotion by means of fuzzy logic. The values of Speed, Interval-Difference and Interval-Variation are the input to the fuzzy system. The following lines explain how these values are obtained. Suppose that we have an eight notes phrase with a length of two beats (see Figure 1).

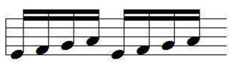

Fig. 1. An example of a phrase.

The Speed is calculated dividing the number of notes in the phrase by its length in beats:

$$Speed = \frac{Notes}{Length} \Rightarrow Speed = \frac{8}{2} \Rightarrow Speed = 2$$

The Interval-Difference, the difference (in intervals) between the first and last notes, is equal to $2\frac{1}{2}$ tones (the first note in the phrase is a E and the last note is A):

$$IntervalDifference = 2.5$$

As explained earlier, the Interval-Variation is the average distance (in intervals) between all the notes within a phrase. In this case, the distance between the first and the second notes is equal to $\frac{1}{2}$ tone, the distance between the second and the third notes is equal to 1 tone, the distance between the third and the fourth notes is equal to 1 tone, and so on:

$$IntervalVariation = \frac{.5 + 1 + 1 + 2.5 + .5 + 1 + 1}{7} = 1.071...$$

Now, we can apply the fuzzy logic (see figure 2). We obtain the following membership values (MV):

Speed Fast with MV = 1.0	Variation Small with MV = 0.428
Difference Small with MV = 0.583	Variation Medium with MV = 0.571
Difference Medium with MV = 0.416	

The minimum and maximum values of the universe of discourse of the variables Speed, Variation and Difference are determined experimentally.

The next step is to apply a set of fuzzy rules (we have defined 54 rules); the implication operator we employ is Mamdani and the aggregation operator is union. For this example, the rules selected for emotion are:

1. IF Speed IS Fast AND Difference IS Small AND Variation IS Small THEN Emotion = Happy
2. IF Speed IS Fast AND Difference IS Small AND Variation IS Medium THEN Emotion = Happy
3. IF Speed IS Fast AND Difference IS Medium AND Variation IS Small THEN Emotion = Happy
4. IF Speed IS Fast AND Difference IS Medium AND Variation IS Medium THEN Emotion = Very Happy

And the rules selected for tension are:

1. IF Speed IS Fast AND Difference IS Small AND Variation IS Small THEN Tension IS Very Calm
2. IF Speed IS Fast AND Difference IS Small AND Variation IS Medium THEN Tension IS Calm
3. IF Speed IS Fast AND Difference IS Medium AND Variation IS Small THEN Tension IS Neutral
4. IF Speed IS Fast AND Difference IS Medium AND Variation IS Medium THEN Tension IS Neutral

Employing the membership values for Speed, Difference, and Variation, we obtain the following results:

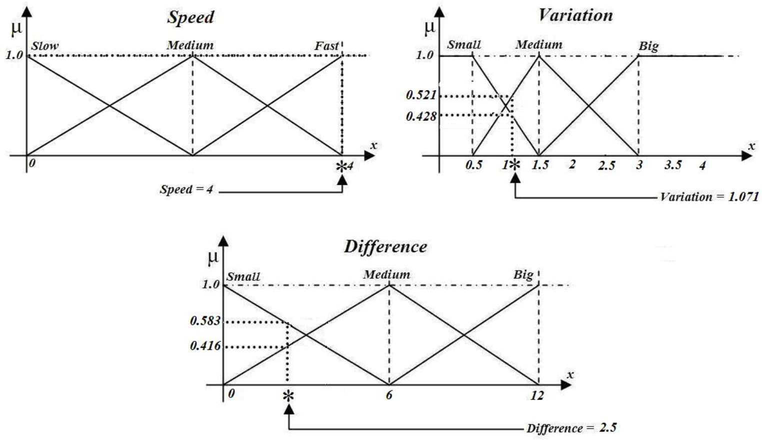

Fig. 2. The values of Speed, Difference and Variation, and their membership values for the fuzzy sets.

Emotion	Tension
Rule1 = Happy with MV = 0.428	Rule1 = Very Calm with MV = 0.428
Rule2 = Happy with MV = 0.571	Rule2 = Calm with MV = 0.571
Rule3 = Happy with MV = 0.416	Rule3 = Neutral with MV = 0.416
Rule4 = Very Happy with MV = 0.416	Rule4 = Neutral with MV = 0.416

The next step is to perform the defuzzification using the Center Of Area (COA) method. So, having in mind the values of the singletons (Happy=80, Very_Happy=100, Very_Calm=20, Calm=40 and Neutral=60) the calculation is performed as follows:

$$Emotion = \frac{(Happy*0.428)+(Happy*0.571)+(Happy*0.416)+(Very\_Happy*0.416)}{0.428+0.571+0.416+0.416}$$

$$Emotion = 84.54$$

And for the tension:

$$Tension = \frac{(Very\_Calm*0.428)+(Calm*0.571)+(Neutral*0.416)+(Neutral*0.416)}{0.428+0.571+0.416+0.416}$$

$$Tension = 44.41$$

Finally, the value of the tension is modified by the average octave of the phrase. This occurs because the higher the octave is the higher the tension that the phrase produces. So, for this example the average octave of the phrase is 4, therefore the tension is equal to 177.64:

2.2 Building the Knowledge Base

In ERMEG a Previous Melody is defined as a sequence of CPs:

$$\text{Previous-Melody} = \text{CP1} + \text{CP2} + \text{CP3} + \ldots \text{CPn}$$

Each CP has associated emotional and tensional values, which we refer to as Local-Emotion and Local-Tension. But from CP2 onwards we can also associated to each phrase a Historical-Emotion and a Historical-Tension. The Historical-Emotion is a vector that records the value of the Local-Emotion of each phrase that has preceded the current CP; the Historical-Tension is a vector that records the value of the Local-Tension of each phrase that has preceded the current CP. In this way, the Historical-Emotion of CP3 is equal to [Local-Emotion of CP1, Local-Emotion of CP2].

All these information is recorded into a structure known as Context; so, the Context is comprised by four elements: the local emotional and tensional values of the current CP, and its historical emotional and tensional vectors. Each time a CP is performed, its emotional and tensional values are calculated and the Context is updated. We refer to the content of the Context after CP1 is performed as Context1; we refer to the content of the Context after CP2 is performed as Context2; and so on (see Figure 3).

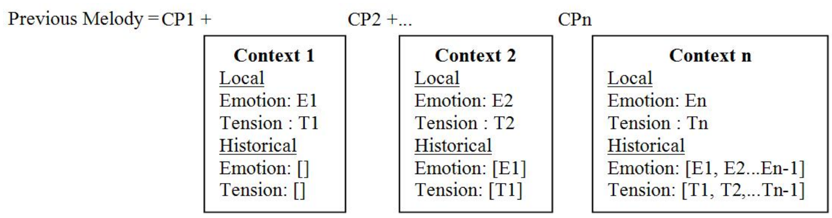

Fig. 3. Representation of a Previous Melody and its Contexts.

Then, the system links contexts with the next CP in the melody. So, Context1 is associated with CP2, Context2 is associated with CP3, Context(n-1) is associated with CPn. The purpose of this type of associations is to establish possible ways to continue a composition given a specific context. In this way, the system records information like "when a melody has a context like Context 1, a possible way to progress the melody is performing CP2". With this information, the system creates in memory its knowledge structures -known as atoms- that it will employ during the engagement-reflection cycle to generate novel melodies.

Thus, in order to create its knowledge structures, the system performs the following process for each Previous Melody.

1. The system initializes the Context and makes the first phrase in the melody the current CP.

2. The system calculates its Emotional and Tensional values and updates the Context.
3. The content of the Context is copied into a new structure known as Atom.
4. The system associates the following CP in the melody to the Atom; we refer to this phrase as the next possible action to be performed in the melody.
5. The system takes the next CP in the melody, makes it the current CP and goes back to step 2 until all phrases in the melody are processed.

3 The Engagement-Reflection Cycle.

ERMEG generates melodies as a result of an engagement-reflection cycle.

3.1 Engagement

The engagement cycle works as follows:

1. The user provides an initial phrase.
2. The system calculates its emotional and tensional values and updates the Context. This is the initial Context.
3. The Context is employed as cue to probe memory. The system looks for all atoms which are equal or similar to the current Context and retrieves their associated CP.
4. The system selects at random one of the CPs retrieved as the next (musical) action to be performed in the melody in progress.
5. The selected CP is appended to the melody in progress, the Context is updated, and the cycle starts again (it goes to step 3).

Engagement ends when an impasse is declared (i.e. no atom can be matched), or when a predefined number of CPs have been added to the melody in progress.

During engagement the system employs two important parameters to decide if a Context matches an Atom (i.e. to decide if the Context is equal or similar to the Atom): the Resolution-Constant and the ACAS-Constant. The Resolution-Constant determines if two values are considered as equivalent. For example, if the Resolution-Constant is set to 90%, the Local-Emotion in the Context is equal to 95 and the Local-Emotion in an Atom is equal to 100, they are considered equivalents. That is, any value between 90 and 100 is considered as equivalent to 100. The ACAS-Constant indicates the minimum percentage that the Context must be equivalent to the Atom in order to produce a match. For example, if the ACAS-Constant is set to 50%, at least the 50% of the Context must be equivalent to the Atom. The values of these constants have important effects during engagement. If the Resolution-Constant is set to 100% and the ACAS-Constant is set to 100%, we are forcing the system to reproduce any of the Previous Melodies. On the other hand, if the Resolution-Constant is set to 1% and the ACAS-Constant is set to 1%, the system is forced to select at random between all the CPs in its knowledge-base the next phrase to continue the melody.

3.2 Reflection

During Reflection the system tries to break impasses and evaluates the coherence and novelty of the melody in progress.

Breaking Impasses. When an impasse is declared, the system decrements in 10% the value of the Resolution-Constant and the ACAS constant, and then switches back to engagement. The purpose is to make easier to match an atom. If after modifying the constants the system cannot match any atom, the system tries a second strategy: it selects at random a CP from its knowledge base, added it to the melody in progress and switches back to engagement. The purpose of this strategy is to modify the Context in a way that now it can match an atom. If the system fails again the impasse is declared as unbreakable and the E-R cycle ends.

Evaluation of Coherence. In ERMEG a melody is coherent when there are not abrupt jumps of notes between two continuous phrases within a melody. For example, let us imagine that the last note of CP1 is the first degree and the first note of CP2 is the seventh degree; in this case, the distance between them is too big ($5\frac{1}{2}$ tones) and therefore the coherence is broken (the maximum allowable distance between two continuous phrases can be modified by the user of the system). To solve this problem the system moves the second phrase closer to the first one, or creates a new CP that works as a bridge that joins both phrases in a smooth way. For example, suppose that engagement generates the two phrases in Figure 4. Reflection evaluates that composition as incoherent. The system modifies the melody either by moving back the second phrase closer to the first one and in this way reducing the distance between them (Figure 5.a), or by building a bridge of notes in two measures that will join the two phrases (Figure 5.b).

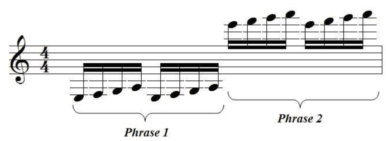

Fig. 4. Two phrases generated by engagement.

Evaluation of Novelty. In this work, a melody is considered original when it is not similar to any of the Previous Melodies. The system compares the melody in progress with all Previous Melodies. If they are similar the system decrements the value of the Resolution and ACAS Constants in order to look for novel atoms to match.

The engagement-reflection cycle ends when an impasse is unbreakable, or when the system performs a predefined number of E-R cycles.

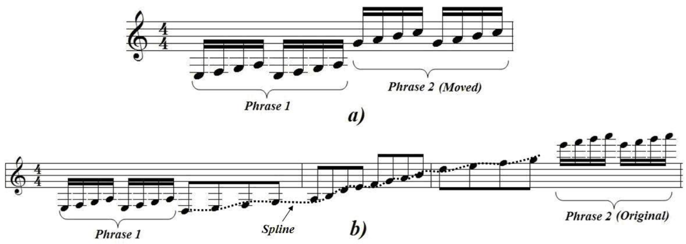

Fig. 5. Modification made by reflection.

4 Discussion.

Figures 6 and 7 show two melodies generated by ERMEG. The system performed the same number of engagement-reflection cycles for each of the examples. The following parameters were employed in both cases:

- Resolution-Constant : 90%
- ACAS-Constant: 50%
- Maximum difference allowed between two consecutive notes in two different phrases: 3 Tones

ERMEG employed the same group of Previous Melodies to generate the two melodies. However, for the melody in Figure 6, the system employed CPs with a length of 8 beats and for the melody in Figure 7 the system employed CPs with a length of 2 beats. In both figures, the first measure corresponds to the initial phrase provided by the user, which was the same for both examples; the shaded parts in the figures show those notes inserted or modified during reflection; the rest are the phrases generated during engagement.

Although the melodies shared the same initial phrase, and the same group of Previous Melodies, they progressed in different ways. There were two main reasons for this behavior: the content of Atoms depended on the size of the CPs; the number of Atoms, and therefore the number of options to progress a melody, also depended on the size of the CPs. Thus, each time we modified the length of the CPs, we altered the content and number of knowledge structures inside the system. In the case of the melody in Figure 6 it was easy to recognise the origin (in the set of Previous Melodies) of each phrase. This made sense because the building blocks, i.e. the CPs, were quite long: 8 beats. However, in the case of the melody generated with CPs equal to 2 beats (Figure 7), it was hard to identify where each phrase came from.

The main purpose of this work was to explore if the ideas that inform the E-R model employed for building MEXICA could be applied for musical composition. ERMEG suggests that the E-R model can be employed to generate novel monophonic melodies.

Music composition is a very complex process. It involves several aspects that have not been taken into consideration in this work. There are several computer

Fig. 6. A melody generated by ERMEG with a composition-phrase equal to 8 beats.

Fig. 7. A melody generated by ERMEG with a composition-phrase equal to 2 beats.

programs that produce very interesting results in this area. We are interested in generating a computer model of creativity that can be applied in different domains. We believe this work is a good first step towards a more complete E-R Model.

References

1. Pérez y Pérez, R & Sharples, M. (2001). MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, 13, 119-139.
2. Gelernter, D. 1994. *The Muse in the Machine*. London: Fourth Estate.
3. Pérez y Pérez, R. (2007). Employing Emotions to Drive Plot Generation in a Computer-Based Storyteller. *Cognitive Systems Research*. Vol. 8, number 2, pp. 89-109.
4. Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, 80(3), B1-B10.
5. Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S. & Peretz, I. (2008) Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology*, vol. 68, pp. 17-26
6. Juslin, P. & Sloboda, J. A. (2001) *Music and Emotion: Theory and Research*. Oxford: Oxford University Press.