THE JOURNAL OF VISUALIZATION AND COMPUTER ANIMATION, VOL. 5: 247–264 (1994)

Automatic Background Music Generation based on Actors' Mood and Motions

JUN-ICHI NAKAMURA, TETSUYA KAKU, KYUNGSIL HYUN, TSUKASA NOMA AND SHO YOSHIDA*
Department of Artificial Intelligence, Kyushu Institute of Technology, 680–4 Kawazu, Iizuka, Fukuoka, 820 Japan

SUMMARY

Since adding background music and sound effects even to short animations is not simple, an automatic music generation system would help improve the total quality of computer generated animations. This paper describes a prototype system which automatically generates background music and sound effects for existing animations. The inputs to the system are music parameters (mood types and musical motifs) and motion parameters for individual scenes of an animation. Music is generated for each scene. The key for a scene is determined by considering the mood type and its degree, and the key of the previous scene. The melody for a scene is generated from the given motifs and the chord progression for the scene which is determined according to appropriate rules. The harmony accompaniment for a scene is selected based on the mood type. The rhythm accompaniment for a scene is selected based on the mood type and tempo. The sound effects for motions are determined according to the characteristics and intensity of the motions. Both the background music and sound effects are generated so that the transitions between scenes are smooth.

KEY WORDS Music generation Actors' mood Actors' action

INTRODUCTION

Background music and sound effects are important to describe actors' mood and movements vividly in animations. Using existing music is not so simple as usually thought. It is not easy to find pieces of music that match given scenes well. To get music that synchronizes with the transitions of given scenes, we may have to hire musicians for particular animations. Considering the cost of this approach, we may consider computer generation of background music, although the quality would not be as good as we want at the moment. Since adding background music and sound effects even to short animations is not simple, an automatic generation system will be a good news for those who want to create quality animation. Although the proposed method can be applied to generate music at the same time as animations are generated, this paper considers only the problem of adding music to existing animations generated by some means.

To present this study, it is appropriate to review the field of automatic music composition. Music composition mechanisms suggested so far can be classified into stochastic (chance) techniques and deterministic techniques. Hiller and Isaacson are regarded as the first to compose music successfully with computers. They composed a composition entitled

* Sho Yoshida is currently in Kyushu Institute of Design, Shiobaru 4-9-1, Minami-ku, Fukuoka 815 Japan.

CCC 1049–8907/94/040247–18 *Received June 1994*

Illiac Suite for string quartet in 1959.[1] Their method can be considered to belong to the stochastic approach. They followed the age-old practice of expressing musical styles as constraints, detailing what is and is not appropriate. They enforced seven melodic constraints and nine harmonic constraints. These constraints for ensemble music were influenced by the conventions of the 18th century counterpoint, but were extremely crude even by elementary pedagogical standards. Then music was generated by chance techniques. For example, pitches were selected using the following mechanism: a pitch was chosen at random from the major scale. If this pitch satisfied all the constraints, the program went on to the next note. Otherwise, a new pitch is chosen at random again, repeating the constraint satisfaction process. At bottom, Hiller's model of composition is rule-satisfaction driven by a random number generator. In such stochastic techniques, the composer has little control over the fine detail.

To alleviate this control problem, in the 1960s Mathews and his colleagues made some experiments with deterministic compositional mechanisms. They investigated pitch quantization as a compositional strategy.[2] In one experiment, they generated a counterpoint between two voices by the following method: choose a pitch for Voice 1, then choose a pitch for Voice 2 so that the interval between them is quantized to the nearest 3rd, 4th, 5th or 6th from Voice 1. Mathews and Rosler represented music as a set of functions of time that could be generated either from keyboards and knobs, or by algorithm.[3] In one experiment, they represented the pitch and rhythmic lines of two folk melodies as functions of time. Then they computed a running interpolation between the two sets of functions in such a way as to produce hybrid melodies. Even though the rules of compositions are deterministic, one does not know what will be the outcomes of their combinations until one actually combines them. The result is sequences that are unplanned in fine detail by the composer. This lack of control over fine detail is still present in deterministic techniques, even though it is less than in chance techniques.

In Europe, Koenig[4,5] developed composing programs Project I and Project II. Project I takes very skeletal musical material, i.e. pitches, loudness and durations, and using a combination of serial and stochastic procedures, produces a musical score. Project II takes detailed specifications of compositional rules in addition to musical material, thus giving a more control over the compositional process. It seems that computer composition has a long way to go to be truly automatic and intelligent.

The techniques that use quantization of pitches, music parameters as functions of time, or serialism are rather challenging to common people including the authors. Some of these techniques are too 'progressive' and some of them require inputs difficult for common people with high school music training to specify. Hiller's approach, which is most conventional, looks like the most manageable. From the viewpoint of animators who want to add background music to a given animation, Hiller's techniques seem to be too random. For example, pitch generation is quite random except that it has to satisfy given rudimentary constraints. For each scene of a given animation, however, animators would like to choose musical motifs, so that they may match well the scene in terms of the perceived mood of the scene, and characteristics of motions in it. Then they would want the rhythm and harmony accompaniment to be generated automatically based on the mood of the scene. Our approach also uses a kind of chance technique in generating chord progression. But rather than choosing a chord transition at random and then checking if it satisfies given constraints, as in Hiller's approach, we assign chord progressions probabilities based on common musical sense, so that chord transitions are generated according to these probabilities.

THE OVERVIEW OF THE MUSIC GENERATION SYSTEM

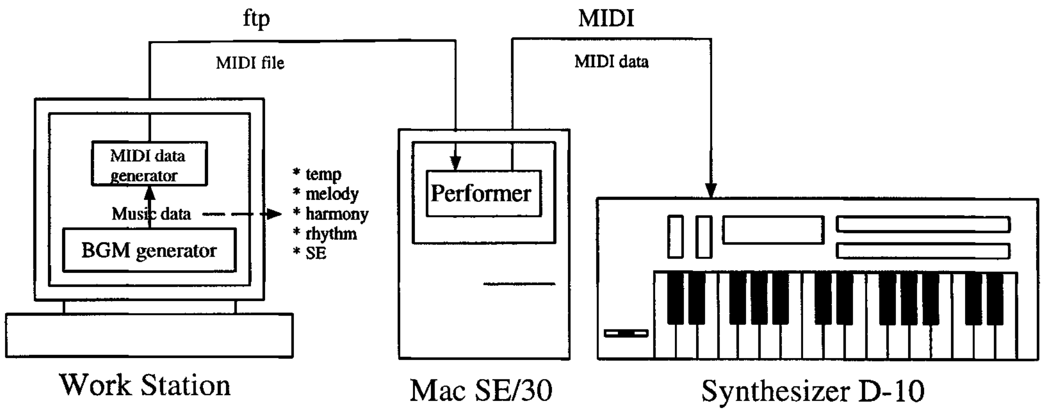

Figure 1. The system diagram. The Background Music generator produces the music data which consists of the tempo, melody, harmony, and rhythm accompaniments, and sound effects for the music. The MIDI data generator simply converts the format of the music data into the form readable by Performer. Performer reads the MIDI file and generates the music signals to control the Synthesizer

The overall structure of the music generation system is shown in Figure 1. The background music generator is implemented in Prolog and the MIDI * data generator is implemented in C on a Unix workstation. Performer is a software package on the Macintosh for controlling digital musical instruments.

Here we briefly describe how the background music for a *given* animation is composed and generated from the point of view of users as follows:

1. Watch the animation and divide it into a sequence of scenes. Divided scenes should have clear transitions between them. Using a stopwatch, measure the length, that is the duration, of each identified scene.
2. Having determined the duration of each scene, specify 'music' and 'motion' parameters. The music parameters specify information needed to generate music for a give scene. The motion parameters describe the actions of the actors that the user is watching. In the following paragraph, we will discuss these parameters more precisely.
3. Submit the input parameters for the scenes to the background music generator which generates a MIDI file as shown in Figure 2.
4. Transfer the MIDI file to the Macintosh Performer. Play the Performer together with the Video Displayer in order to replay the animation with the background music and sound effects.

The music parameters are used to generate background music. For each scene, they consist of the following items (Figure 2):

1. *Scene label*: the sequence number of a scene.
2. *Motive sequence*: a sequence of motifs for the scene. A motif consists of only one

* MIDI stands for Musical Instrument Digital Interface, which is a popular interface for digital musical instruments.
A motif is the smallest unit of melody in music.

```
[1,                                      % scene label
 [[a,1],[b,1]],                          % motives sequence
 [sleepy,2],                             % mood type
 16.9,                                   % length of the scene
 [m(acou_piano_1),h(harpsi_1),
  r(elec_bass_1)]]                       % timbre
```

Figure 2. An example of music parameter. This example specifies scene 1 as follows. We use two motifs: `[a,1]` *and* `[b,1]`. `a` *and* `b` *refer to basic motifs, and the numbers refer to their variations in the motif database. The main actor is a little sleepy as denoted by* `[sleepy,2]`. *This scene continues for 16·9 s. Three timbres are for playing*

measure* in the present system. This is the basis of background music for the scene. We may assign a specific motif sequence to each actor in the animation. If the motif sequence for a scene is the one assigned to the main actor of the scene, the background music emphasizes his appearance.

3. *Mood type*: the mood of the main actor in the scene. In the current implementation, it is one of 'glad', 'happy', 'sleepy', 'sad', 'angry' and 'tired' with its degree (1–5). When actual music is generated from the motif sequence, it is influenced by the mood type of the scene.
4. *Length of the scene*: the duration of the scene. This is used to adjust the tempo.
5. *Timbre*: the timbre is the 'colour' of the sound (e.g. piano-like, flute-like or violin-like), which depends on the synthesizer. Although this should be determined from the main actor and his mood type, we have no definite information to compute this. Therefore we manually specify it.

The motion parameters of a scene are simply the actions that constitute the scene. The system generates sound effects characterizing the actions of actors, if there are natural sound effects for them as in falling and hitting. They consist of actors' names, actions, their degrees (1–5), the start-times, and the end-times. An example of a motion parameter is shown in Figure 2.

The relations among a scene, its mood type and motion parameters are shown in Figure 4. In this example, the scene consists of a sequence of actions, that is sleeping, falling and hitting. The mood type *sleep* for scene 1 corresponds to the mood of the main action, which is Action 1 (sleeping). Although one scene may contains several actions of the main actor, a scene is assigned only one mood in the current system, which is the mood of the primary action. We need to find a method to combine several moods of the main actor for one scene. If we divide a given animation into a sequence of scenes so that one scene contains only one action, this problem seems to go away. But then most of the scenes would be too short to give 'well-formed' music. How to resolve this potential conflict is a future research problem.

THE PROCESS OF MUSIC GENERATION

We describe the steps of music generation and show how the input music and motion parameters of Figures 2 and 3 are transformed into background music and sound effects.

* Measures are basic units in music. Each measure is marked by the preceding and following vertical bars in the standard music notation.

```
[1,                          % scene label
 [[rabbit,                   % actor
   falling, 3                % action and its degree
   15.6, 16.2]]]             % start time and end time
```

Figure 3. An example of a motion parameter. This example specifies that the actor, the rabbit, falls with a medium speed as denoted by the intensity 3 in the scene 1. This motion starts at 15·6 s from the beginning of the scene, and ends at 16·2 s.

Given the input music and motion parameters for a scene, the system produces the music data by the following steps, as shown in Figure 5:

1. Determine the tempo and the number of measures for the scene by using the mood type and its degree. For instance, the mood type 'very sad' makes the basic tempo 'very slow.' The basic tempo is represented by the number of quarter notes in one minute. Determine the number of measures to fill the scene by using the tempo and the length of the scene.
2. Determine the scale and the volume of the music for the scene. The scale, e.g. major or minor, and the volume also can change the impression effectively. For a 'sad' scene, the minor scale and a low level of volume are selected, for instance.
3. Generate a chord * progression by using harmonic progression rules and the number of the measures of the scene.
4. Generate a melody for the scene by repeating and transforming the motif for the scene based on the chord progression and the number of measures of the scene. Although each motif is a short melody which does not have a enough length for an arbitrary scene, repeating and transforming them enables us to generate a melody of an arbitrary length.
5. Generate an accompaniment, which is composed of harmony and rhythm accompaniments, by retrieving them from the accompaniment database with the mood type as a key.[6] Even if the same melody is used, the impression of the background music is changed when different accompaniments are used. When 'happy' harmony and rhythm accompaniment are used, the scene becomes 'happy'. We have 151 harmony accompaniments and 163 rhythm accompaniments in the accompaniment database.
6. Generate sound effects depending on the actors' motions and adjust them to match the background music. The sound effect patterns for various motions, such as 'hitting', 'blinking', and 'falling', are stored in the database. These patterns are modified by the degrees (intensity) of motions. 'Quick falling' is expressed by a fast changing of the notes. Also when some conditions are met, the background music is suspended while the sound effects are generated. It yields a more effective impression. The background music is resumed when the sound effects are terminated.

In the rest of this paper, we discuss the details of the steps.

The tempo and number of measures for a scene

To generate a chord progression and melody for a scene, the number of measures to fill the scene should first be computed. This in turn requires computing the tempo of the scene.

* A chord is three or more musical notes sounded simultaneously.

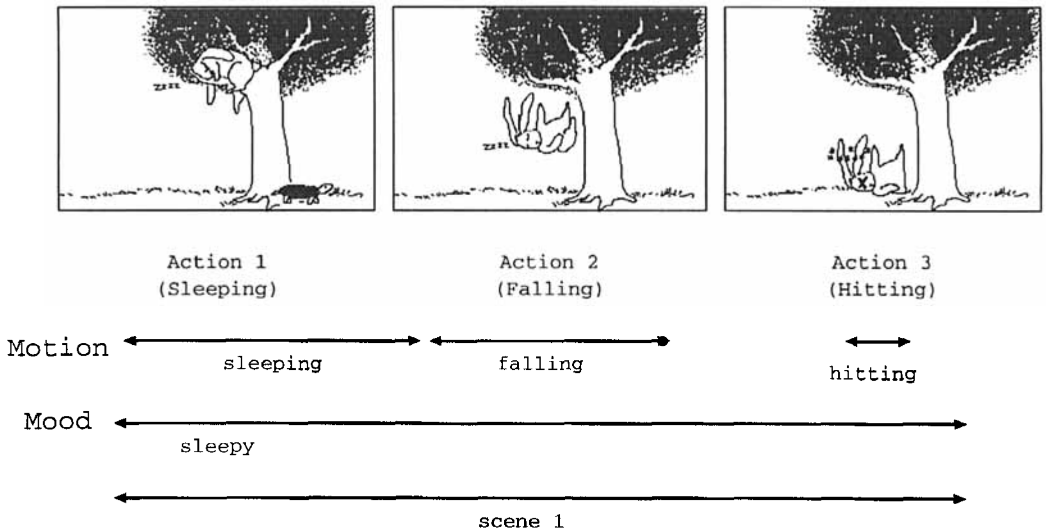

Figure 4. The relations among a scene, its mood type and motions

The basic tempo depending on mood types

To determine the tempo of a scene, we consider the mood type and its duration. We first determine the basic tempo, and then adjust it so that the number of measures may be a multiple of two and the transition to the following scene may be smooth.

Table I. Constants for each mood type

Mood type	t_0	d
glad	100	10
happy	110	8
sleepy	100	−8
sad	110	−10
angry	100	10
tired	90	6

The basic tempo t, which is the number of quarter notes in one minute, is calculated by the expression:

$$t = kd + t_0$$

where k is the degree of the mood type, d and t_0 are constants that depend on the mood type as shown in Table I. This table is decided by experiments. For example, if the mood type is 'glad' and its degree is 1, then the basic tempo becomes 110. If the degree is 5, then the tempo becomes 150. This means that the basic tempo becomes faster according to the degree of the mood type. On the other hand, the basic tempo becomes slower for the mood type 'sleepy' and 'sad.'

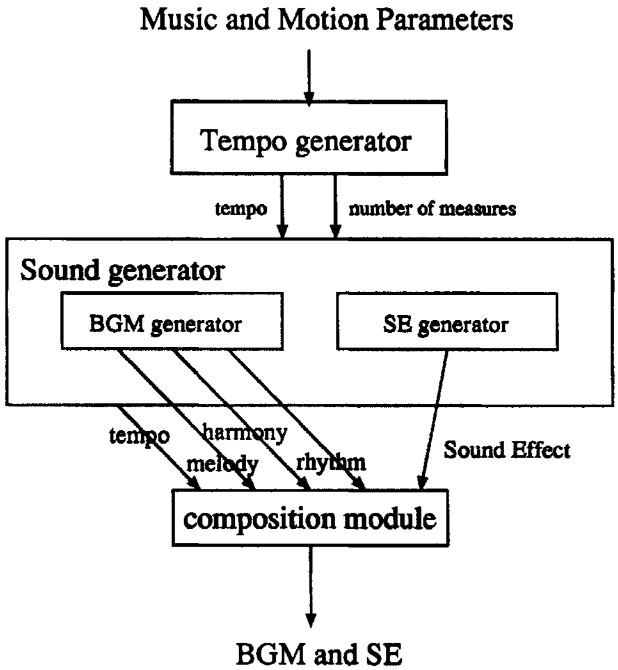

Figure 5. The process of music generation

The number of measures and the adjustment of the tempo

The number of measures m for a scene is calculated as

$$m = \frac{T}{60}\frac{t}{H}$$

Here T is the duration of the scene in seconds ($T/60$ in minutes), t the basic tempo, and H the metre. According to common practices in music, we require that the number of measures of a scene be a multiple of two. So, in the above formula, t should be adjusted so that m may become a multiple of two. Let the adjusted number of measures and the adjusted tempo be

$$m' = 2\lfloor \frac{m}{2} + 0 \cdot 5 \rfloor$$

and

$$t' = \frac{60}{T} H m'$$

respectively. Note that it is more preferable that the number of measures is a multiple of four. Therefore the system first computes the adjusted tempo by $4\lfloor m/4 + 0 \cdot 5 \rfloor$. If the adjusted tempo t' is between 70 and 160, which is 'normal', then the system uses this. If not, the system adjusts m to be a multiple of two.

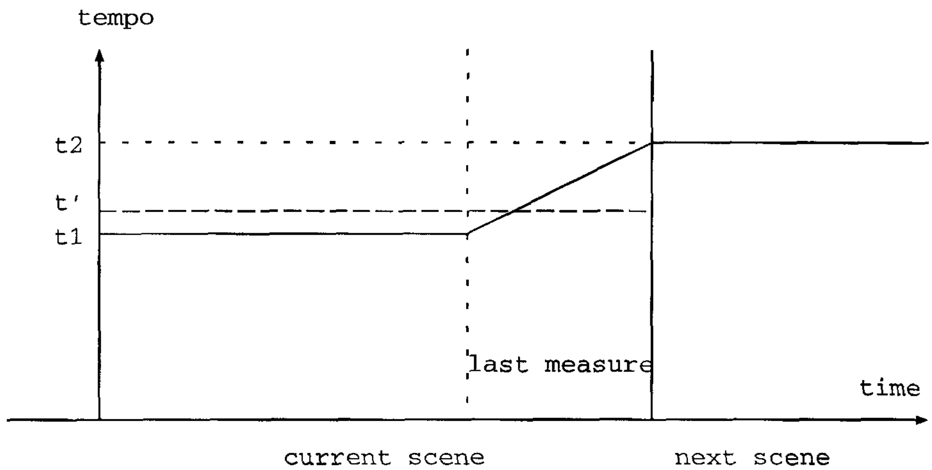

Figure 6. Smooth transition of the tempo from one scene to the next scene

However the adjusted tempo t' is obtained without considering the tempo of the next scene, and may cause a sudden tempo change between the successive scenes. To connect the scenes smoothly, the tempo of the scene should be gradually changed to the tempo of the next scene, as shown in Figure 6. Let the tempo of the current scene be t_1 from the first measure to the beginning of the last measure, and the tempo of the next scene t_2. Then the tempo of the last measure is obtained by linearly interpolating between t_1 and t_2. The new adjusted tempo t_1 is calculated by the equation

$$t_1 = \frac{1 - 2m'}{(1/t_2) - (T/30H)}$$

The details are discussed by Kaku.[7,8]

The generation of background music

The background music for a scene is generated by using the given motifs, the mood type for the scene, and its degree. In the system, the background music consists of three parts: a melody, harmony accompaniment and rhythm accompaniment. The system first generates a chord progression for a scene, and then transforms the given motifs into a melody based on the chord progression. Then, the system chooses a harmony accompaniment and rhythm accompaniment based on the mood and its degree for the scene. The pitches of the harmony accompaniments, and the rhythm accompaniments are arranged based on the chord progression.

The determination of scale and volume

The scale of a music, e.g. major or minor, affects its impression. The system chooses the major scale for the mood types 'glad', 'happy' and 'sleepy', and the minor scale for 'sad', 'angry' and 'tired.' In addition to this, the system determines a key for each scene based

Table II. The key of mood types

Mood type	Degree	Key
happy	1–3	C major
	4–5	G major
glad	1–3	C major
	4–5	G major
sleepy	1–5	C major
sad	1–3	a minor
	4–5	e minor
angry	1–3	a minor
	4–5	e minor
tired	1–5	C major

on the mood type and its degree. Moreover, the key of the current scene is determined by considering the key of the previous scene, in order to make the two keys related.* If the adjacent keys are not related, the music sounds strange. Given the key of the previous scene, the current scene can have one of the related keys. One of the related keys is chosen based on the mood type and degree of the current scene, as shown in Table II. For example, for a 'glad' or 'happy' scene with degree 5, the key is modulated to G major from the previous key. For a 'sad' scene with degree 5, the key is modulated to E minor from the previous key. An example of key modulations is shown in Table III.

The volume of the background music is also adjusted according to the mood type. The system increases the volume for 'glad', 'happy' and 'angry', and decreases the volume for 'sleepy', 'sad' and 'tired'. This is also derived from the general intuition for each mood type.

The generation of chord progression

In general, when composers create music, they consider melody and chord progression simultaneously, so that they may match each other. However, doing the same thing automatically is difficult. So, this system first generates a chord progression for a scene, and then transforms the given motifs into a melody based on the chord progression. Each measure has only one chord associated with it, and chords change every measure. For example, if a scene needs eight measures, it will have eight chords.

Important kinds of chords used in music include triads, 7th chords, 9th chords, chromatic chords, and their inversions. Among them, triads, their first inversions, and the dominant 7th chord are used most frequently in music.† In this system, only these three kinds of

* To be related, keys should have harmonic relationship. For example, G major and D major are closely related, because D major is the dominant of G major, and these two keys differ only by one sharp.

† A triad chord consists of three notes such that the highest note is fifth higher than the lowest note, and the second highest note is third higher than the lowest. The first inversion of a triad is made by moving the lowest note up by one octave. The first inversion is notated by adding the suffix $_6$ to the chord name, as in I_6. The dominant chord is a triad whose lowest note

Table III. An example of key modulations

Scene No.	Mood type	Degree	Key
1	sleepy	5	G major
2	angry	5	G major
3	tired	5	C major
4	glad	5	G major
5	happy	5	G major
6	glad	5	G minor
7	fun	5	G minor

chords are used. Examples of these chords are shown in Figure 7. However, the chords VI_6 and VII are excluded because they are not frequently used in music.

Table IV. The states of chord progressions

States	Chords
start	I
middle end	V V_7
end1	II IV
end2	V V_7
end3	I
normal	I I_6 II II_6 III IV IV_6 V V_6 VI VII VII_6

The system generates chord progression based on harmonic rules. The harmonic rules are based on those most commonly used in classical music of the 17th to 19th centuries. There are no absolute harmonic rules in music. However, there are some favoured ways of using chords in music. According to common practices, in this system, the music of each scene starts with chord I, and ends with the perfect cadence‡, as shown in the finite state automata of Figure 8. If one scene has more than eight measures, the middle of the melody ends with the half cadence.§ Note that in the chord transition automaton, the states 'normal', 'middle end', 'end1' and 'end2' contain more than one chord, as shown in Table IV. A chord progression is generated by following the state transitions of the automaton. When a transition occurs to a state that contains more than one chord, one of them should be chosen. For example, when a transition occurs from the state 'start' to the state 'normal', one out of the nine chords in the state 'normal' should be chosen. This decision is also made based on common practices. For example, we can find the chord progression 'V–I' in many classical pieces, but we cannot find the chord progression 'V–IV' as much. The probabilities of chord transitions that are believed to reflect common practices in music are codified in Table V. The chord transition probabilities are determined by a standard harmony textbook (Reference 9, p. 20) and an author's sense of music. For example, the probability of moving from chord II to chord V is 0·3. The probability of moving from

is the fifth note of a major or minor scale. The dominant 7th chord is made by adding another note above the dominant chord so that that note is seventh higher than the lowest. For example, the dominant 7th chord of C major consists of notes G,B,D,F.

‡ The perfect cadence is one of the most popular ways of ending a melody. The perfect cadence ends with a chord progress of V to I.

§ The half cadence ends with chord V.

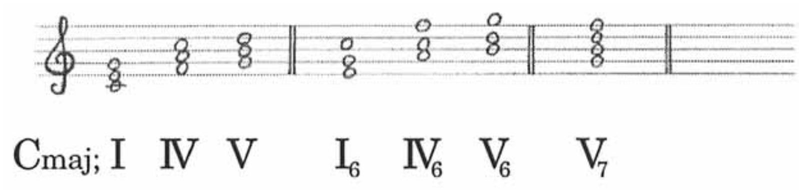

Figure 7. Examples of chords

chord V$_6$ to chord I is 1·00. This means that chord V$_6$ always moves to chord I.

Table V. The probabilities of chord transitions

From	To (Probabilities: 100 means probability 1)
I	I$_6$(5), II$_6$(10), III(5), IV(25), IV$_6$(5), V(25), V$_6$(5), V$_7$(5), VI(10), VII$_6$(5)
I$_6$	I(40), IV(20), IV$_6$(20), V(10), V$_7$(10)
II	I(5), II$_6$(10), IV(10), IV$_6$(10), V(30), V$_7$(15), VI(10), VII$_6$(10)
II$_6$	I(20), IV(20), V(50), V$_7$(10)
III	IV(10), IV$_6$(30), VI(6O)
IV	I(30), II(10), II$_6$(10), V(40), V$_7$(10)
IV$_6$	I(50), VII(50)
V	I(70), VI(20), VII$_6$(10)
V$_6$	I(100)
V$_7$	I(30), VI(70)
VI	I(10), II(50), III(10), IV(10), IV$_6$(10), V(5), V$_6$(10), V$_7$(5)
VII$_6$	I(60), I$_6$(40)

Figure 9 shows a set of possible eight-measure chord progressions that are generated for the input motifs of Figure 2. In this system, chord progression is not dependent on any input parameter. The chord progression for each scene is generated purely according to the chord transition probabilities, starting with chord I.

The generation of melody

Once a chord progression is determined for a scene, a melody for the scene is generated from the given motifs according to the chord progression. In this system the length of one motif is one measure. A total of seven motifs are provided in the motif database. The motifs for a scene are chosen by users. Two motifs out of the seven are assigned to a scene. The system repeats and varies the two motifs of a scene so that they can fill the whole scene. The melody of a scene consists of one or two phrases, * depending on the number of measures of the melody. If the melody has four or six measures, it will be one phrase. If the melody has more than eight measures, it consists of two phrases. For example, consider the scene of Figure 2 which is assigned two one-measure motifs *a* and *b*. Suppose that the

* A phrase is an intermediate unit of melody. In the Classical period, it typically has four measures. Since it usually ends with some form of cadence, it is also the unit of a harmonic progression.

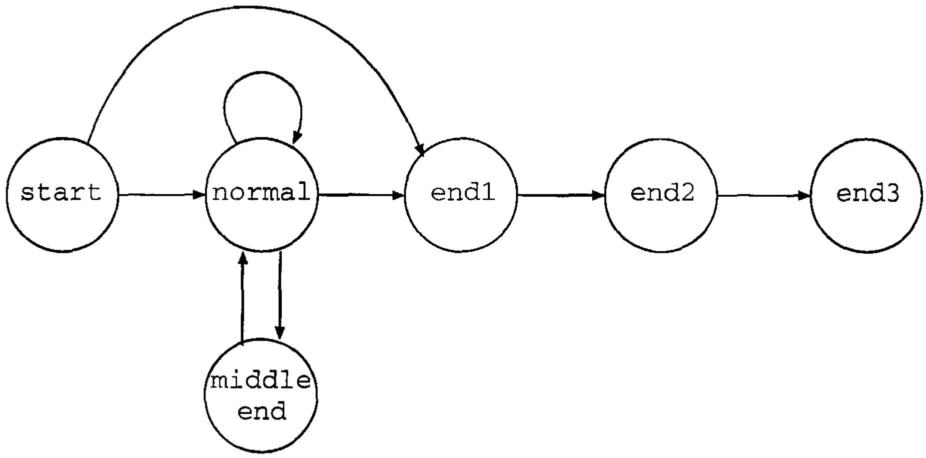

Figure 8. The diagram of chord progressions

scene requires eight measures. Then, a phrase is generated by repeating and varying the first motif a twice according to the chord progression, and then by repeating and varying the second motif b two time, as shown in Figure 10. Then the second phrase is generated by varying the first phrase according to the chord progression.

$\mathrm{I} - \mathrm{II}_6 - \mathrm{I} - \mathrm{V} - \mathrm{I} - \mathrm{IV} - \mathrm{V}_7 - \mathrm{I}$
$\mathrm{I} - \mathrm{II}_6 - \mathrm{I} - \mathrm{V} - \mathrm{I} - \mathrm{IV} - \mathrm{V}_7 - \mathrm{I}$
$\mathrm{II} - \mathrm{VII}_6 - \mathrm{I} - \mathrm{V} - \mathrm{VI} - \mathrm{II} - \mathrm{V}_7 - \mathrm{I}$
$\mathrm{I} - \mathrm{II}_6 - \mathrm{I} - \mathrm{V} - \mathrm{VI} - \mathrm{IV} - \mathrm{V}_7 - \mathrm{I}$
$\mathrm{I} - \mathrm{V} - \mathrm{I} - \mathrm{V}_7 - \mathrm{I} - \mathrm{IV} - \mathrm{V} - \mathrm{I}$
$\mathrm{I} - \mathrm{IV} - \mathrm{I} - \mathrm{V} - \mathrm{I} - \mathrm{IV} - \mathrm{V}_7 - \mathrm{I}$

Figure 9. Examples of chord progressions

$$\underbrace{\text{a1, a2, b1, b2,}}_{\text{the first phrase}}\ \underbrace{\text{a3, a4, b3, b4.}}_{\text{the second phrase}}$$

Figure 10. Transforming and repeating motifs

The variation of a motif according to a chord progression means moving the notes of the motif in parallel by the 'distance' between the adjacent chords in the chord progression. Since each motif is designed to have only one chord, it can be easily transformed to other chords and other keys. The motifs used in the example of Figure 2 are shown in Figure 11.

The melody of a scene starts with a down beat.* To make the melody smooth, the last

* A beat is the smallest unit within a measure. For example, a quarter note is one beat in metre 4/4. One measure usually has four beats in metre 4/4. A down beat is the first beat of a measure.

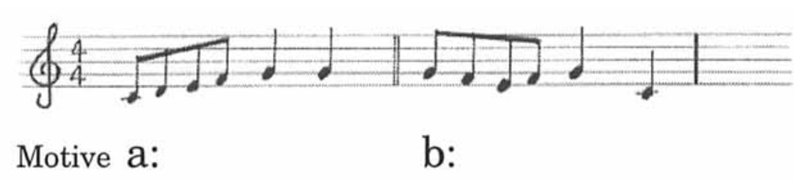

Figure 11. Example motifs

measure of the first half of the melody consists of an one long note * determined according to the chord of the measure. And the last measure of the melody consists of a long tonic note which is the first note of the scale. The range of a melody used in the system is from G2 to A4.

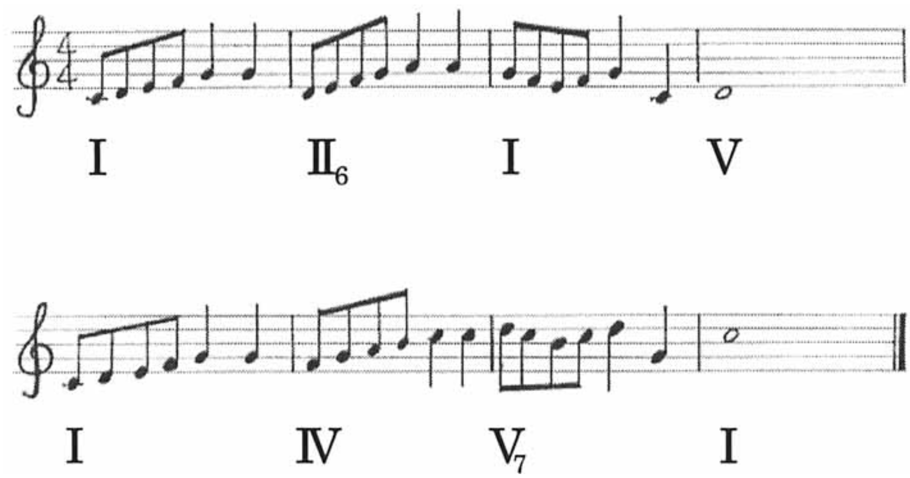

Figure 12. An example of melody generation

For example, the system generates a melody as shown in Figure 12 from the motifs of Figure 11, according to the chord progression of Figure 9.

The generation of harmony accompaniments

To enrich music, harmonic accompaniment is added to the melody in this system. Harmonic accompaniments are various ways of playing chords. Examples of harmonic accompaniments are shown Figure 13.

The harmonic accompaniment for a scene is retrieved from the accompaniment database using the mood type as a key. The 151 harmonic accompaniments taken from the textbooks by Takeuchi and Iwama[10] and Suzuki[11] are classified according to the six mood types ('glad', 'happy', 'sleepy', 'sad', 'angry' and 'tired') and their degrees (1–5). This classification is

* A long note is the whole note in the case of metre 4/4.

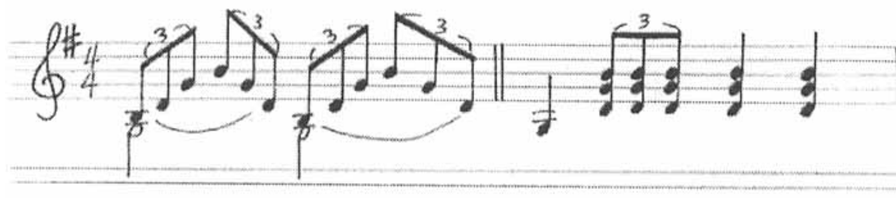

Figure 13. Examples of harmony accompaniment

done by considering the range of pitch, complexity of rhythm, the number of notes and harmony.[6] A selected harmonic accompaniment for the melody of Figure 12 is shown Figure 15.

Figure 14. Examples of rhythm accompaniment

The generation of rhythm accompaniments

In this system, rhythm accompaniment is also added the melody to enrich the music. Rhythm accompaniments are various kinds of rhythm patterns which play the base notes of chords over and over again. The rhythm accompaniment has no relationships with the rhythm of the melody. In this system, rhythm accompaniments play the role of base melody in classical music. Examples of rhythm accompaniments are shown Figure 14.

The rhythm accompaniment for a scene is also retrieved from the accompaniment database. The 163 rhythm accompaniments taken from the handbook[12] are classified according to two aspects: speed ('fast', 'normal' and 'slow') and complexity ('complex', 'normal' and 'simple'). The speed type for each rhythm accompaniment is determined by the number of its notes. A rhythm accompaniment that consists of many notes is classified as 'fast'. The complexity type is determined by the number of different notes in each rhythm accompaniment. An accompaniment that consists of various notes is classified as 'complex'.

The rhythm accompaniment is selected according to Tables VI and VII by the basic tempo and the mood type. A selected rhythm accompaniment for the melody of Figure 12 is shown Figure 15.

The generation of sound effects

Sound effects are used to emphasize actors' motions. Most motions generate their own sounds. If there is no sound in a scene, it may be strange. Some motions such as entering and blinking, do not generate sound. However, figurative sound can be used to emphasize

Figure 15. Harmony and rhythm accompaniments of a melody

them. Sound effects are classified into sound generated by musical instruments and sound generated by objects. The former are mainly figurative. The latter are mainly real sounds, such as the noise of an explosion. Since the current synthesizer of the generation system can produce only sounds of musical instruments, the authors have paid attention to figurative sound effects. By observing several existing animations, we found that figurative sound effects are used in many scenes.

Table VI. Basic tempo and its speed

		Tempo			Speed
		Tempo	$\leq$	110	slow
110	$<$	Tempo	$\leq$	130	normal
130	$<$	Tempo			fast

Table VII. Mood types and complexity

Mood type	Complexity
glad, angry	complex
happy, sad	normal
sleepy, tired	simple

Sound effect patterns

The sound effects generated by musical instruments are classified as follows:

1. Sounds without scale (generated by percussion)
   (a) momentary movements (e.g. hitting, knocking): single percussion.
   (b) continuous movements (e.g. earthquake, preparation): continuous percussion.
2. Sounds with scale (generated by wind or string instruments)
   (a) momentary movements (e.g. entering, blinking): short melody
   (b) continuous movements (e.g. falling, rising): chromatic scale.

Table VIII. Sound effect patterns

Percussion		Scale instruments	
Motion	Type	Motion	Type
Hitting	Single	Surprised 1	Short melody
Boxing	Single	Surprised 2	Short melody
Knocking	Single	Blinking	Short melody
Explosion	Single	Shocked	Short melody
Preparation	Continuous	Nodding up	Short melody
Headache	Continuous	Nodding down	Short melody
Earthquake	Continuous	Entering	Short melody
Numbed	Continuous	Falling	Chromatic scale
Hand Clapping	Continuous	Rising	Chromatic scale

The 18 sound effect patterns shown in Table VIII are used. Each pattern is defined by

1. The motion name.
2. The type of sound or pattern of melody.
3. The base parameters of volume, length and type of musical instrument.
4. The constants for the modification.

Note that 'surprised' and 'shocked' are treated as 'motions' whose sound effects are generated using short melodies. These, however, may be described as the mood types. Since the present system assumes that all moods are static enough to generate a long piece of music, it cannot treat short ones. Therefore short moods are represented by sound effects rather than by music, as if they are action with sound effects. This is not logical, because it makes moods and actions mixed up. This problem will be solved in a future system.

The system selects a pattern according to the input motion parameter, and modifies it by the degree of motion. The volume and tempo are also adjusted by the degree.

Table IX. Motions and their degrees needed to stop the background music

Motion	Degree	Motion	Degree
Preparation	1	Surprised	3
Headache	3	Shocked	2
Numbed	3	Entering	1
Hand clapping	4	Falling	1
		Rising	1

The interaction of sound effects with the background music

In some cases, it is more effective to stop the background music when generating sound effects. By experiments, we found motions and their degrees that require the background music to be stopped during the generation of sound effects, as shown in Table IX.

The system stops the background music exactly at the beginning of the sound effects, even if it is the middle of a note. It produces more effective sound effects. The system resumes the background music starting from the new note just after the sound effects. This produces a more natural transition.

EXAMPLES

The background music and sound effects are generated for the advertisement video of the authors' department (about four minutes), and for the animation called *Tom & Jerry* (about two minutes). The background music and some of the sound effects for the advertisement video were adequate. The transition between scenes was smooth. However, many of the sound effects were strange. This is because the advertisement video does not usually use figurative sound effects, and the sound effect patterns used were those of the Tom and Jerry. The sound effects for the Tom and Jerry were adequate, though the background music was poorer than the original one. Since this evaluation heavily depends on our intuition, quantitative evaluation of the results is future research topic.

CONCLUSIONS

This study is a new attempt to incorporate background music and figurative sound effects automatically into existing animations by means of music synthesizers. The background music of the scene is generated from the motifs assigned to each scene by various music composition rules, according to the mood of the main actor. Hiring musicians to play background music for particular animations is expensive. Hence, we submit that the generation of background music by means of a music synthesizer is useful. Because the users can easily play simple motifs by means of synthesizers, they can examine the outcome of background music as they change the motifs. Throughout this paper, however, we assumed that we were trying to add music to *existing* animations. The problem of generating music together with animation from a unified specification is challenging and is a future research topic.

To improve the quality of music in this system, there are some problems to be addressed. The system uses a limited number of chords, so that chord progressions are too simple. So we need more variety in them. The melody generated by moving the motifs in parallel according to the chord progression, may not sound smooth. So we need to gather the common constraints on melody progressions and generate melodies to satisfy those constraints. In this system, melody generation depends on the chord progression not the other way around. But to be realistic, chord progression and melody should influence each other. Appropriate harmony and rhythm accompaniments are chosen by retrieving them from the accompaniment database with the mood types as keys. But they need to be chosen more intelligently. This system generates only a style of solo music with accompaniments. It is desirable to make ensemble music, for example string quartet or orchestra. We need to provide a greater number of motifs, and a better interface in which the users can find appropriate motifs for scenes. Finally, the authors are planning to connect this system with an automatic animation generation from a natural language story.[13] This project will clarify the usefulness and problems of the present system.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the large contributions of Keiko Mori to prepare the accompaniment databases, and to thank Moon R. Jung for reading the paper critically and giving constructive suggestions.

REFERENCES

1. L. Hiller and L. Isaacson, *Experimental music*, McGraw-Hill Book Company, Inc., New York, 1959.
2. M. V. Mathews, 'The digital computers as a musical instrument', *Science*, **142**, (1963).
3. M. V. Mathews and F. R. Mooer, 'GROOVE—a program to compose, store, and edit functions of time', *Communications of the ACM*, **13**, (12), (1970).
4. G. M. Koenig, 'Project one', *Electronic Music Reports 2*, Institude of Sonology, Utrecht, 1970.
5. G. M. Koenig, 'Project two', *Electronic Music Reports 3*, Institude of Sonology, Utrecht, 1979.
6. K. Mori, 'Generation system of accompaniment for automatic arrangement', *Graduation Thesis*, Department of AI, Kyushu Institute of Technology, 1992, in Japanese.
7. T. Kaku, T. Noma and J. Nakamura, 'An experimental study of BGM generation with emotional parameter', *Proc. 6th Convention*, Japanese Society of Artificial Intelligence, 1992, pp. 719–722, in Japanese.
8. T. Kaku, 'Automatic BGM generation based on actor's emotion and movements', *Master Thesis*, Department of AI, Kyushu Institute of Technology, 1993, in Japanese.
9. W. Piston, *Harmony*, TAELIM CHULPANSHA, Seoul, Korea, 1977, in Korean.
10. T. Takeuchi and M. Iwama, *Accompaniment*, Yamaha ONGAKU-SHINKOU-KAI, 1976, in Japanese.
11. A. Suzuki, *Introduction for Accompaniment on Piano Playing*, AYUMI-SHUPPAN, 1988, in Japanese.
12. Y. Shoya, *Rock Dram Handbook*, SHINKO-music, 1988, in Japanese.
13. T. Noma, K. Kai, J. Nakamura and N. Okada, 'Translating from natural language story to computer animation', *First Singapore International Conference on Intelligent Systems*, Raffles Convention City, Singapore, 1992.