

An Interactive and Concerted Dance System ~ Emotion Extraction and Support for Emotional Concert ~

Ikuko Harada[†] Makoto Tadenuma[†] Takahiro Nakai[†] Ryotaro Suzuki[†]
Naoto Hikawa[†] Mao Makino[†] Masayuki Inoue^{†‡}

[†] ATR Media Integration & Communications Research Laboratories

E-mail: {harada, tadenuma, tnakai, ryotaro, hikawa, mao}@mic.ATR.co.jp

[‡] Hiroshima Institute of Technology

Department of Information and Intellectual Systems Engineering

E-mail: minoue@cc.it-hiroshima.ac.jp

Abstract

A human activity that expresses one's emotion is significantly important in non-verbal communications. Human body motion such as a dance performance is one of the important factors in expressing such an emotion. In this paper, an effective method to extract emotional information in real time from dance images is described. A novel heuristic function is also proposed to evaluate the degree of concert between two dancers. As a typical application, a dance system called MIDAS that interactively combines dancers' images with video, sound, and CG characters is introduced. MIDAS can visualize how the dancers are in good concert with each other and encourage their expressions.

1. Introduction

Many communication technologies including the telephone, e-mail, and even web pages, etc. have been devoted to verbal communications. Verbal information plays important roles in logical communications, but non-verbal information brings a lot of emotional information and constitutes another important side of human communications.

We are aiming at enhancing such non-verbal communications and at establishing its framework utilizing multimedia technologies. Art is an appropriate reference for this research, because artists express their emotions by creating artwork. Some related works can be seen from this point of view. For example, Tanaka et al. proposed a method utilizing composition information in pictorial art [9] and Suzuki et al. discussed video art synthesis [8].

An emotional expression through a dance performance is a good case study for non-verbal communication research,

because dance is a form of art on which human motion, which is also an important and indispensable factor in non-verbal communications, is focused. Therefore, human motion analysis is of great interest in this research.

As for an emotion analysis from human motion, some research results have been reported. In particular, Prof. Camurri and his group have been working on an analysis of dance movements and have developed some applications [1]. They introduced a method to characterize human motion with emotions by tracing human body movements and by abstracting limb behaviors as changes in bounding rectangles. We adopt a similar approach to abstract a dancer's behaviors. We use the acceleration of the extracted parameters as additional information and properly categorize the emotions in real time according to the theory of choreography.

Because supporting and enhancing a dancer's emotive expressions as well as analyzing his/her emotions are our ultimate goals, a dance system called MIDAS (MIC Interactive Dance System; MIC stands for our Media Integration & Communication Research Labs.) that interactively encourages the dance performance is being developed. The emotion extraction method and MIDAS system were first introduced by Y. Iwadata et al. [4].

The emotions from one dancer can be extracted and MIDAS can interact with these emotions. However, many dances are performed by two or more dancers. In such dance performances, not only the emotions of each dancer but also their cooperativeness is thought to be one of the critical features. Accordingly, we introduce the concept of emotional concert to evaluate how movements of dancers are concerted.

In this paper, the degree of emotional concert is defined as a heuristic criterion and an extension of interactive dance system MIDAS is proposed to visualize how dancers are

well concerted with each other.

In the following sections, choreographic considerations are first introduced. Then, the emotion extraction method we propose is briefly discussed. Next, the degree of emotional concert is defined. Then, the extended MIDAS system that utilizes both dancers' emotions and the degree of emotional concert is explained. Next, some considerations about MIDAS and future work are discussed. Finally, a conclusion summarizes this paper.

2. Overview of dance research

Choreographers have been long studying human motions in order to understand their physical features and to effectively express emotional images through body motions. The basis of the human body motion theory in modern dance was established by Rudolf Laban (1879-1958). Laban proposed three kinds of descriptions for human motions: Motif description, Effort-shape description, and Structural description [2]. The Motif description provides the most salient feature of a motion. The Effort-Shape description can describe movements in terms of quality and expression. The Structural description is well known as Labanotation. These descriptions are used to systematically record movements. One of the typical examples of Laban's theory can be seen in robotics [6].

The Effort-Shape description refers to the two concepts of "Effort" and "Shape." The "Effort" concept in particular is appropriate for analyzing emotional representations and consists of four elements: space, weight, time, and flow. These elements define the qualities of movements. Time has the categories of sudden movements and sustained movements. Space stands for a design of the posture and is the directivity of the posture and movement. Weight stands for the strength and power of the movement. Weight is categorized into firm movements and fine touch movements. Flow stands for the carefulness versus easiness that can be seen in a movement. Among these four elements, time, space, and weight are known as more fundamental than flow.

Moreover, a related study of Laban's theory has shown that emotional images of dance can be categorized into seven typical motives [5, 7]. In this research, the three elements of time, space (mentioned as design in [5]), and energy are used for the categorization. These elements are the equivalents of time, space, and weight in Laban's theory, respectively. The seven motives represent seven kinds of emotions: happy, solemn, lonely, natural, sharp, dynamic, and flowing. Table 1 shows the relationship between the Effort-Shape description and the seven motives.

The Effort-Shape description and the seven motives provide the background knowledge to analyze the emotions represented in dance movements.

Table 1. Relationship between seven motives and Time-Space-Energy.

Motives	Time	Space	Energy
Happy	speedy		light
Flowing	smoothing	curving	
Lonely	slowly		sustained
Natural	regular	balanced	sustained
Solemn	gradual		massive, sustained
Sharp	accented	straight	firm, sudden
Dynamic	speedy, accented	spread	light

3. Emotions extraction from human motions

The concepts of the Effort-Shape description and the seven motives are thought to be useful for analyzing dancers' movements. However, these concepts are still rather emotions-based. Accordingly, the effort elements should be defined strictly with parameters able to be extracted from sequences of dance images. Moreover, the purpose of the emotion analysis should be to determine a function that maps these elements to one of the seven motives.

3.1. Image processing

Given a video sequence of a dance performance, the human motion parameters should be extracted in real time as an abstraction that features the dancer's emotional information. We adopt the center of gravity and a bounding rectangle of a dancer's silhouette image as the abstraction features as shown in Figure 1. The movements of the center of gravity represent the dancer's movements on a dancing stage. The bounding rectangle represents the dancer's local motions by limb behaviors [3].

Let W_n be the ratio of the areas of the silhouette and the bounding rectangle, and let X_n and Y_n be defined as the X and Y coordinates of the center of gravity in the n th video frame. We adopt a physical parameter vector \mathbf{x}_n , defined by the following equation, to extract "Time," "Space," and "Energy:"

$$\mathbf{x}_n \stackrel{\text{def}}{=} (W_n, \dot{W}_n, \ddot{W}_n, |(\dot{X}_n, \dot{Y}_n)|, |(\ddot{X}_n, \ddot{Y}_n)|)^T, \quad (1)$$

while \dot{a}_n and \ddot{a}_n are time differentials calculated by $\dot{a}_n = a_n - a_{n-1}$ and $\ddot{a}_n = a_n - 2 \cdot a_{n-1} + a_{n-2}$, and $|(a, b)|$ is obtained by $\sqrt{a^2 + b^2}$.

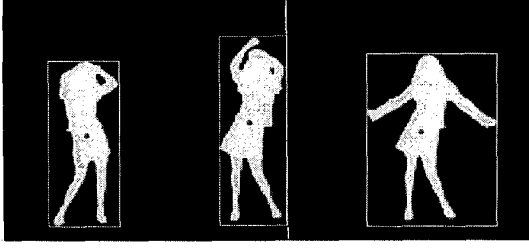


Figure 1. Silhouette Images of a Reference Dancer.

3.2. Time, space, and energy

It seems that the "Time," "Space," and "Energy" elements can be defined as a direct correspondence with physical parameters like a moving speed, a silhouette area ratio, and a moving acceleration. But, because the "Effort" elements are rather the emotional representations of movements, therefore the correspondence are to be better determined through psychological experiments.

Thus, we define the correspondence using SD(Semantic Differential) method. Video scenes of a professional dancer were used for an emotion analysis. The dancer performed in accordance with the above-mentioned seven motives. Then, subjects answered the strength of their emotional impressions for each video scene about previously prepared key words. The principal component analysis extracted the dominant factors from the those evaluations. The three dominant principal components could be thought as the "Time," "Space," and "Energy" elements, respectively.

Let a vector $\mathbf{P} (= (P_1, P_2, P_3)^T)$ be the component scores of the answered strengths of one video scene for the dominant three components. Assuming those scores can be obtained by the linear equation of the physical parameters defined by the equation (1), the following equation is obtained:

$$\mathbf{P} = A\mathbf{x} + \mathbf{b}, \quad (2)$$

while A is the 3×6 coefficient matrix, \mathbf{b} is the constant factor vector, and \mathbf{x} is the average of physical parameters defined by the equation (1) in the video scene. The A and \mathbf{b} can be obtained applying the multiple regression method.

3.3. Mapping to seven motives

We need to clarify the relationship between Time-Space-Energy and the seven motives. Therefore other psychological experiments were carried out to categorize reference dance sequences. In these experiments, the subjects vote for one motive out of the seven motives for each reference dance sequence. We assume that the strengths of the seven

motives in each dance scene can be represented by linear combinations of time, space, and energy as follows:

$$V_k = \mathbf{c}_k^T \mathbf{P} + d_k, \quad (3)$$

while V_k represents each of the seven motive values and $k = \{1, 2, \dots, 7\}$. The multiple regression method is applied to find the appropriate coefficient vector \mathbf{c}_k and the constant d_k . The motive decision is made by selecting the motive having the maximum V_k value.

4. Emotional concert

As described in Section 3.3, the motive of a dancer can be extracted from the dancer's silhouette for each video frame. When a few or more dancers make a dance performance, the dancing motive of each dancer can be known by analyzing the motive independently. Then, one of the next important factors of the dance is to understand how the dancers are cooperatively dancing with one another.

In this section, the degree of concert in the sense of emotions is considered. The degree of emotional concert is defined from the motive values of both of two dancers in a performance. For simplicity, only the degree of concert for two dancers is considered.

4.1. Motive profile

If two dancers are dancing with the same motives, the dancing of both dancers can be expected to be rather concerted. However, this criterion is too naive to determine the cooperativeness of dancers. Using seven motive values obtained by equation (3), the motive profile \mathbf{M} can be defined as the sequence of those scores:

$$\mathbf{M} \stackrel{\text{def}}{=} (V_1, V_2, \dots, V_7) \quad (4)$$

Let $V(\mathbf{M}, k)$ be the i th value of motive profile \mathbf{M} , i.e., V_k in Equation (3) ($k = \{1, 2, \dots, 7\}$). The average value \bar{M} and the variance $\sigma_{\mathbf{M}}$ of profile \mathbf{M} are obtained as follows:

$$\bar{M} = \frac{1}{7} \sum_{k=1}^7 V(\mathbf{M}, k) \quad (5)$$

$$\sigma_{\mathbf{M}} = \frac{1}{7} \sum_{k=1}^7 (V(\mathbf{M}, k) - \bar{M})^2 \quad (6)$$

Let a normalized profile $f_N(\mathbf{M})$ be a transform of the original profile \mathbf{M} such that the resultant average and variance are transformed into 0 and 1. The motive value of $f_N(\mathbf{M})$ is obtained for each $k (= \{1, 2, \dots, 7\})$ as follows:

$$V(f_N(\mathbf{M}), k) = \frac{1}{\sqrt{\sigma_{\mathbf{M}}}} (V(\mathbf{M}, k) - \bar{M}) \quad (7)$$

These values can be thought to characterize the emotional information precisely in order to define the degree of emotional concert as a heuristic criterion that evaluates the difference in dance performances in the emotional sense.

4.2. Degree of emotional concert

The degree of emotional concert $C(\mathbf{M}_1, \mathbf{M}_2)$ is defined using two motive profiles \mathbf{M}_1 and \mathbf{M}_2 . In order to evaluate the degree of concert, three factors are considered. The factors are obtained from normalized profiles, variants of profiles, and averages of profiles, in order.

The first factor $C_P(\mathbf{M}_1, \mathbf{M}_2)$ is a square of the difference of normalized profiles:

$$C_P(\mathbf{M}_1, \mathbf{M}_2) = \sum_{k=1}^7 (V(f_N(\mathbf{M}_1), k) - V(f_N(\mathbf{M}_2), k))^2 \quad (8)$$

This factor is thought to be the most important factor because the normalized profiles are thought to characterize so to speak a blending ratio of motives included in the dance.

The second factor $C_V(\mathbf{M}_1, \mathbf{M}_2)$ is obtained from variants of profiles:

$$C_V(\mathbf{M}_1, \mathbf{M}_2) = \frac{1}{2} \left(\frac{\sigma_{\mathbf{M}_1}}{\sigma_{\mathbf{M}_2}} + \frac{\sigma_{\mathbf{M}_2}}{\sigma_{\mathbf{M}_1}} \right) - 1 \quad (9)$$

This factor has a minimum zero when two variants are the same and monotonically grows when their ratio goes far from equality. It compares the amplitudes of the profiles, in other words, the clarities of the motive.

The third factor $C_A(\mathbf{M}_1, \mathbf{M}_2)$ is a comparison of average motive values weighted by standard deviations as follows:

$$C_A(\mathbf{M}_1, \mathbf{M}_2) = \left| \frac{\overline{M}_1}{\sqrt{\sigma_{\mathbf{M}_1}}} - \frac{\overline{M}_2}{\sqrt{\sigma_{\mathbf{M}_2}}} \right| \quad (10)$$

This factor is thought to compare the total emotional strengths of performances.

The degree of concert is defined as a linear combination of the above three factors:

$$C(\mathbf{M}_1, \mathbf{M}_2) \stackrel{\text{def}}{=} w_1 C_P(\mathbf{M}_1, \mathbf{M}_2) + w_2 C_V(\mathbf{M}_1, \mathbf{M}_2) + w_3 C_A(\mathbf{M}_1, \mathbf{M}_2), \quad (11)$$

where w_1 , w_2 , and w_3 are the weights of three factors, which are experimentally determined.

The degree of concert is always a non-negative value. The smaller the value becomes, the higher the cooperativeness of the dance gets.

5. MIDAS system

An interactive dance system called MIDAS is described in this section. This system utilizes emotion extraction and

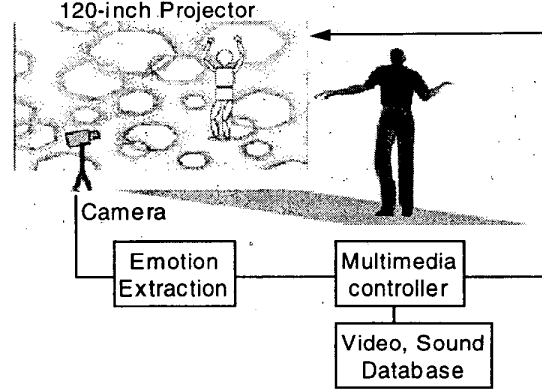


Figure 2. Schematic of MIDAS Base System.

the degree of concert calculation functions in the above sections.

5.1. MIDAS base system

The base system of MIDAS extracts the emotions of a user, and then it expresses the emotions through a multimedia controller. In this system, the extracted emotional information is sent to a multimedia controller. The multimedia controller manages a video switcher, real-time disk system, and sound system. The multimedia controller interprets the received emotional information and selects adequate video and sound clips.

The selected video clips are displayed on a 120-inch projection monitor. The selected sound clips are replayed simultaneously. Figure 2 shows a schematic of the base system. All of the contents are designed by graphic artists and a sound creator along the emotional characteristics of each motive [10].

A performer's image is synthesized on the video clips. The time, space, and energy parameters are used to make the performer's image. Figure 3 shows an appearance of the MIDAS base system.

5.2. MIDAS with emotional concert

The MIDAS system described in Section 5.1 can be extended for two dancers using the degree of concert. MIDAS can exhibit the two dancers' performance with multimedia devices according to their own motives. When their dance comes to be concerted in the emotional sense, MIDAS changes its exhibitions according to the degree of concert.

5.2.1. System configuration. MIDAS has two independent cameras and emotion extraction units, which extract motive profiles in the same way as described in Section 5.1.

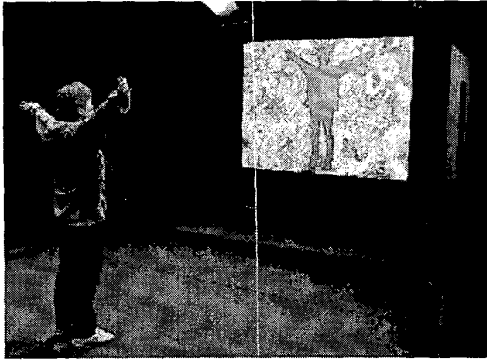


Figure 3. Appearance of MIDAS Base System.

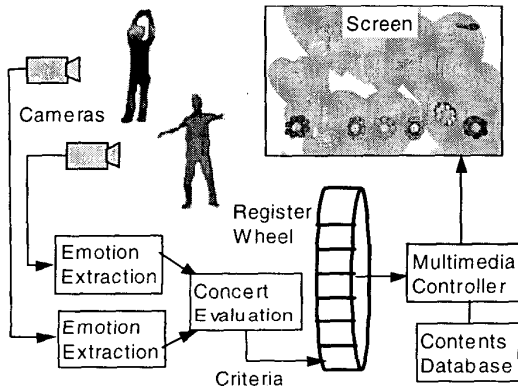


Figure 4. Schematic of Concerted MIDAS.

The concert evaluation unit calculates the degrees of concert according to Equation (11) from the motive profiles obtained from both emotion extraction units in a constant interval time. The degrees of concert and motives of the dancers are stored in a register wheel. According to the moving average of the degrees and most frequently detected motives, a multimedia controller determines the exhibitions to be seen on the screen. Figure 4 shows a schematic diagram of MIDAS with emotional concert.

5.2.2. Contents and exhibitions. Background video clips, 3D CG characters, and sound contents are designed for each motive. One of these contents is selected for each dancer. Therefore, two contents are simultaneously shown on the screen. Figure 5 shows a screen shot when the "Happy" and "Sharp" motives are selected; "Happy" is for the left-side person and "Sharp" is the right-side person. Sound tracks are divided for melody and base parts. Two speakers are used. The first speaker is used for the melody of the motive of the first dancer, and the second speaker is

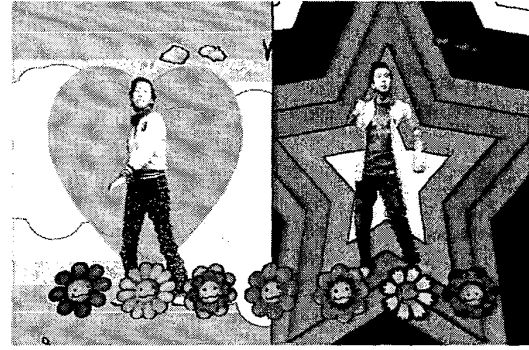


Figure 5. Two Dancers' Appearance.

Table 2. Variety of Contents for Each Grade.

Grade	Dancer Image	Screen	CG Characters
Best	Silhouette with Best Effect	Merged	Cheering with Special Effect
Good	Silhouette with Good Effect	Merged	Cheering
Normal	Silhouette with Normal Effect	Divided	Natural
Bad	Raw Image	Divided	Dull

for the base of the motive of the second dancer.

The video and CG contents have four grades according to the degree of concert. Each grade has a name: "Best," "Good," "Normal," and "Bad." Table 2 shows the variety of contents.

First, for the bad grade, the screen is divided and the contents for each dancer are independently displayed in each divided part of the screen. Raw images of the dancers are overlaid on the video clip. When they get better in terms of cooperation, the screen is merged in one part; the dancers' images, for the best grade, are changed into silhouettes with the best effects and cheering characters move with special effects. An example of the best concert for a "Dynamic" dance is shown in Figure 6. MIDAS outputs a cheering sound and video when it detects still better concert in the best grade to notify their excellence in the emotional co-operations for the current motive. Therefore, dancers can enjoy a best concerted dance for every motive.

6. Considerations on MIDAS and future works

In our experiences, we found that many persons who tried dancing in MIDAS were often puzzled at first by the image and sound outputs from MIDAS, because the outputs easily changed when they tried to follow the outputs. Indeed, the outputs of MIDAS always follow the dancers'

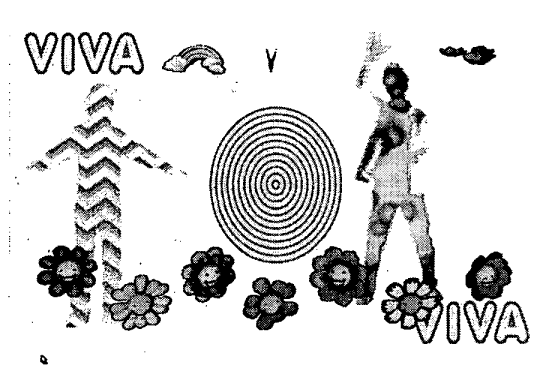


Figure 6. Best Emotional Concert for a "Dynamic" Motive.

movements. User interviews showed that users did enjoy changing images through MIDAS. They seemed to easily learn how to control or how to interact with MIDAS and were able to enjoy expressing their own images through MIDAS.

As a result, MIDAS seems to have immersive effects to its users, because users can see themselves immediately on the screen with many effects and because they do not need to be bothered by wearing any sensors or markers.

The degree of concert proposed in this paper amplifies another aspect of dance performances. When two dancers are cooperatively dancing, they can observe improving grades of contents. This game like effect can promote the communications between the dancers. Getting the best grades for all motives is a great motivation for dancing.

The MIDAS technology is expected to be applied to psychotherapy or educational programs. The emotional concert does not require the same motions as dancing. This feature suggests the possibility of encouraging aged people or small children to enjoy their emotion expressions.

The current MIDAS system is designed for two dancers, but the number of people can be extended and MIDAS can come to be a large networked amusement system in the future. The concert of the total dance and a better video data delivery method should be considered as future works.

The emotion extraction methods are expected to be improved through reference dance revisions and refinements to physical parameters. The heuristic parameter used in calculating the degree of concert, the weights of parameters, and the value ranges of contents grades should be improved through subjective experiments.

7. Conclusions

This paper described an emotion extraction method and proposed a method to evaluate the emotional concert of

two dancers. An interactive dance system called MIDAS was also described. Emotional information was categorized according to choreographic research results. The data extracted from images were used to map the images into these categories. Psychological experiments determined the extraction and mapping functions. MIDAS analyzed the dancers' emotions, evaluated their degree of concert, and output video combined with dance and sound clips. MIDAS and its technologies displayed the novel methods for non-verbal human communications. MIDAS will be extended as an emotional multimedia communication system in the future.

Acknowledgments

We greatly appreciate Prof. Mariko Shiba and Ms. Mamiko Sakata of Kobe University for their useful information about the theory of dance and their help in preparing the dancers' video scenes. We also wish to thank Mr. Yousuke Kita and Mr. Tsutomu Kanegae for their cooperation in the system implementation.

References

- [1] A. Camurri, S. Hashimoto, K. Suzuki, and R. Trocca. Kansei Analysis of Dance Performance. In *Proc. IEEE SMC'99*, IV, pages 327–332, 1999.
- [2] A. Hutchinson. *Labanotation*. Dance Books, 1996.
- [3] Y. Iwade. Study on image expression. *IEICE Technical Report*, HCS99-53:87–94, Oct. 1999 (In Japanese).
- [4] Y. Iwade, M. Inoue, R. Suzuki, N. Hikawa, M. Makino, and Y. Kanemoto. MIC Interactive dance system ~ An emotional interaction system ~. In *Proc. 4th International Conf. on Knowledge-Based Intelligent Eng. Systems & Allied Tech.*, pages 95–98, Aug. 2000.
- [5] C. Matsumoto. Dance Research: Problem Situation and Learning of Problem Solving II — Qualities of Movements and Feeling Values. *Proc. Japanese Association of Physical Education and Sports for Girls and Women*, pages 53–89, 1987 (In Japanese).
- [6] T. Nakata, T. Sato, and T. Mori. Expression of Emotion and Intention by Robot Body Movement. In *IAS-5*, pages 352–359. IOS Press, June 1998.
- [7] M. Shiba. Extraction of Kansei Information in Dance Movements. In *Proc. ATR Workshop on Virtual Communication Environments*, pages 70–85, 1998.
- [8] R. Suzuki and Y. Iwade. Multimedia Montage — Counterpoint synthesis of movies—. In *Proc. IEEE ICMC'99*, pages 433–438, 1999.
- [9] S. Tanaka, J. Kurumizawa, A. Plante, Y. Iwade, and S. Inokuchi. Image Re-Composer: A Post Production Tool Using Composition Information of Pictures. In *Proc. IEEE ICMC'99*, pages 439–444, 1999.
- [10] <http://www.mic.atr.co.jp/organization/dept3/index.html>.