Visualization of Music Impression in Facial Expression to Represent Emotion

Takafumi Nakanishi **Takashi Kitagawa**

Graduate School of Systems and Information Engineering
University of Tsukuba,
Tsukuba, Ibaraki 305-8573, Japan
Email: takafumi@mma.cs.tsukuba.ac.jp
takashi@cs.tsukuba.ac.jp

Abstract

In this paper, we propose a visualization method of music impression in facial expression to represent emotion. We apply facial expression to represent the complicated and mixed emotions. This method can generate facial expression corresponding to impressions of music data by measurement of relationship between each basic emotion for facial expression and impressions extracted from music data. The feature of this method is a realization of an integration between music data and the facial expression that convey various emotions effectively. One of the important issues is a realization of communication media corresponding to human Kansei with less difficulty for a user. Facial expression can express complicated emotions with which various emotions are mixed. Assuming that an integration between existing mediadata and facial expression is possible, visualization corresponding to human Kansei with less difficulty realized for a user.

Keywords: Mediadata, Facial Expression, Music Data, Impression, Kansei.

1 Introduction

A large amount of information resources have been distributed in wide area networks. In this environment, a current interface, for example, computer keystrokes, is difficult of computer manipulation for human being. One of the important issues is a realization of communication media corresponding to human Kansei with less difficulty for a user. The concept of "Kansei" includes several meanings on sensitive recognition, such as "impression," "human senses," "feelings," "sensitivity," "psychological reaction" and "physiological reaction."

Generally, it is important to understand each other emotion correctly in our communication. In particular, facial expression is important as media which convey various emotions effectively. The facial expression is one of nonverbal behaviors. The facial expression can express complicated emotions with which various emotions are mixed which cannot be expressed with words.

The researches which realize composition and recognition of the facial expression are done actively. In these researches, Facial Action Coding System (FACS)(Ekman & Friesen 1978, Ekman & Friesen 1987) is used strictly and most widely. FACS describes facial expression with the combination of some Action Units (AU's). AU's are the minimum units of facial expression operations which are visually discernible. These research results have shown the combination of AU's for expressing 6 basic emotions, which are "happiness", "surprise", "fear", "anger", "disgust", and "sadness". The combination of these basic emotions can express complicated facial expression.

There are the followings as previous researches on construction of facial expression, research which mounts AU and creates the picture of expression(Choi, Harashima, & Takebe 1990), research of facial imitation by 3D face robot agent(Hara, & Kobayashi 1996), etc. These researches realize a construction of facial expression which is close to an actual expression.

In this paper, we propose a visualization method of music impression in facial expression to represent emotion. We apply facial expression to represent the complicated and mixed emotions. This method can generate facial expression corresponding to impressions of music data by measurement of relationship between each basic emotion for facial expression and impressions extracted from music data.

We have already proposed a semantic associative search method based on a mathematical model of meaning(Kitagawa & Kiyoki 1993, Kiyoki, Kitagawa & Hayama 1994). This model is applied to extract semantically related words by giving context words. This model can measure the relation between each word, mediadata, and so on.

In addition, we have already proposed a media-lexicon transformation operator for music data(Kitagawa & Kiyoki 2001, Kitagawa, Nakanishi & Kiyoki 2004). This operator can extract metadata which represents the impression of music data as weighted words.

This proposal method can generate facial expression corresponding to impression of music data utilizing the mathematical model of meaning and the media-lexicon transformation operator for music data. The feature of this method is a realization of an integration between existing mediadata, that is music data, and nonverbal behaviors that convey various emotions effectively, that is facial expression. Namely, the purpose of this method is different from the conventional methods.

The system using facial expression can express impressions of mediadata more appropriately compared with the system only using the information such as words. The facial expression can express complicated emotions. Assuming that complicated emotion can be expressed, human and the system can share mutual emotion and the interface corresponding to human Kansei can be realized.

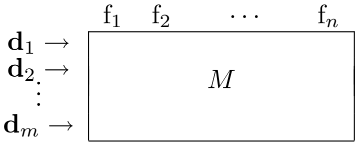
f$_1$
f$_2$
$\cdots$
f$_n$
$\mathbf{d}_1 \rightarrow$
$\mathbf{d}_2 \rightarrow$
$\vdots$
$\mathbf{d}_m \rightarrow$
M

Figure 1: Representation of metadata items by matrix M

2 Mathematical Model of Meaning

The mathematical model of meaning(Kitagawa & Kiyoki 1993, Kiyoki, Kitagawa & Hayama 1994) provides semantic functions for computing specific meanings of words which are used for retrieving mediadata unambiguously and dynamically. The main feature of this model is that the semantic associative search is performed in the orthogonal semantic space. For details, see references (Kitagawa & Kiyoki 1993, Kiyoki, Kitagawa & Hayama 1994).

The mathematical model of meaning consists of:

1. Creation of a metadata space $\mathcal{MDS}$
   Create an orthonormal space for mapping the mediadata represented by vectors (hereafter, this space is referred to as the metadata space $\mathcal{MDS}$). The specific procedure is shown below.

   When m data items for space creation are given, each data item is characterized by n features $(f_1, f_2, \cdots, f_n)$. For given $\mathbf{d}_i (i = 1, \cdots, m)$, the data matrix M (Figure 1) is defined as the $m \times n$ matrix whose i-th row is $\mathbf{d}_i$. Then, each column of the matrix is normalized by the 2-norm in order to create the matrix M.

   (a) The correlation matrix $M^T M$ of M is computed, where M^T represents the transpose of M.

   (b) The eigenvalue decomposition of $M^T M$ is computed.

   $$M^T M = Q \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_\nu & & \\ & & & 0 & \\ & & & & \ddots 0 \end{pmatrix} Q^T, \quad (1)$$

   $0 \le \nu \le n.$

   The orthogonal matrix Q is defined by

   $$Q = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n) \quad (2)$$

   where $\mathbf{q}_i$'s are the normalized eigenvectors of $M^T M$. We call the eigenvectors "semantic elements" hereafter. Here, all the eigenvalues are real and all the eigenvectors are mutually orthogonal because the matrix $M^T M$ is symmetric.

   (c) Defining the metadata space $\mathcal{MDS}$

   $$\mathcal{MDS} := span(\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_\nu). \quad (3)$$

   which is a linear space generated by linear combinations of $\{\mathbf{q}_1, \cdots, \mathbf{q}_\nu\}$. We note that $\{\mathbf{q}_1, \cdots, \mathbf{q}_\nu\}$ is an orthonormal basis of $\mathcal{MDS}$.

2. Representation of mediadata in n-dimensional vectors
   Each mediadata is represented in the n-dimensional vector whose elements correspond to n features. The specific procedure is shown below.

   A metadata for mediadata P is represented in t weighted impression words $\mathbf{o}_1, \mathbf{o}_2, \cdots, \mathbf{o}_t$. These impression words are extracted from media-lexicon transformation operator shown in section 3.

   $$P = \{\mathbf{o}_1, \mathbf{o}_2, \cdots, \mathbf{o}_t\}. \quad (4)$$

   Each impression word is defined as an n dimensional vector by using the same features as the features of the data matrix M.

   $$\mathbf{o}_i = (f_{i1}, f_{i2}, \cdots, f_{in}) \quad (5)$$

   The weighted impression words $\mathbf{o}_1, \mathbf{o}_2, \cdots, \mathbf{o}_t$ are composed to form the mediadata vector, which is represented as an n dimensional vector. The Kansei operator shown in subsection 6 of section 4.2 realizes this composition. The mediadata is represented as mediadata vector which is n dimensional vector by using same features as the features of the data matrix M.

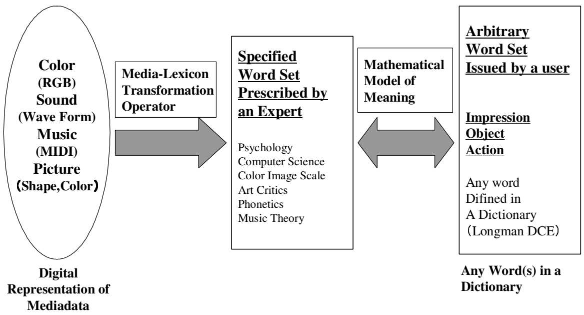
Color
(RGB)
Sound
(Wave Form)
Music
(MIDI)
Picture
(Shape,Color)
Digital Representation of Mediadata
Media-Lexicon Transformation Operator
Specified Word Set Prescribed by an Expert
Psychology
Computer Science
Color Image Scale
Art Critics
Phonetics
Music Theory
Mathematical Model of Meaning
Arbitrary Word Set Issued by a user
Impression
Object
Action
Any word
Difined in
A Dictionary
(Longman DCE)
Any Word(s) in a Dictionary

Figure 2: A framework of media-lexicon transformation operator.

3. Mapping a mediadata vector into the metadata space $\mathcal{MDS}$
   A mediadata vector which is represented in n-dimensional vectors is mapped into the metadata space $\mathcal{MDS}$ by computing the Fourier expansion for a mediadata vector and semantic elements.

4. Semantic associative search
   A set of all the projections from the metadata space $\mathcal{MDS}$ to the invariant subspaces (eigenspaces) is defined. Each subspace represents a phase of meaning and it corresponds to a context. A subspace of the metadata space $\mathcal{MDS}$ is selected according to the context. An association of a mediadata is measured in the selected subspace.

3 Media-lexicon Transformation Operator

In this section, we introduce a media-lexicon transformation operator $\mathcal{ML}$(Kitagawa & Kiyoki 2001).

3.1 A Framework of Media-lexicon Transformation Operator

In Figure 2, we show a framework of the media-lexicon transformation operator $\mathcal{ML}$(Kitagawa & Kiyoki 2001).

$\mathcal{ML}$ is an operator which represents a relation between mediadata and some group of word sets given by a research work by an expert of a specific disciplinary area. The operator ML is defined as

$\mathcal{ML}(Md) : Md \mapsto Ws$

where, Md is an expression of mediadata and Ws is a specific set of words or a collection of word sets usually with weights. The mediadata Md is a specific expression of the mediadata usually in a digital format. The word set Ws is selected by an expert to express impression of the specific media.

By this operator $\mathcal{ML}$, we can search or retrieve the mediadata by arbitrary words issued as a query, using the mathematical model of meaning(Kitagawa & Kiyoki 1993, Kiyoki, Kitagawa & Hayama 1994) which relates any given words to certain word groups dependent on the given context.

3.2 Media-lexicon Transformation Operator for Music Data

Media-lexicon transformation operator for music data(Kitagawa & Kiyoki 2001, Kitagawa, Nakanishi & Kiyoki 2004) extracts some impression words from music data. This operator extracts impression words of a song from elements (musical elements) that determine the form or structure of the song such as harmony, melody, and so on. The fundamental psychological research that examined correlation relationships between impressions and musical elements was conducted by Hevner(Hevner 1935, Hevner 1936, Hevner 1937, Umemoto.ed. 1966). This operator uses the correlation relationships indicated by Hevner to calculate correlations between these sets of musical elements and impression words.

3.2.1 Research of Hevner

In Hevner's research(Hevner 1935, Hevner 1936, Hevner 1937, Umemoto.ed. 1966), key, tempo, pitch, rhythm, harmony, and melody were given as musical elements. Hevner examined the correlation relationships between these 6 musical elements and 8 categories of impression words (Figure 3). Each category of impression words was created by collecting together impression words that had similarities to other words in that category. The 8 categories of impression words were further arranged in a circle so that categories were adjacent to other categories to which they had similarities. Hevner experimentally obtained correlation relationships between musical elements and impressions represented by categories of impression words.

3.2.2 An Implementation Method of Media-lexicon Transformation Operator for Music Data

This section shows an implementation method of media-lexicon transformation operator for music data. For details, see references (Kitagawa & Kiyoki 2001, Kitagawa, Nakanishi & Kiyoki 2004).

This operator consists of three steps:

Step 1 : Composition of Transformation Matrix T.
Transformation Matrix T shown in Figure 4 is composed by the correlation between musical elements and each impression which are given by Hevner.

Step 2 : Extraction of musical element vector $\mathbf{s}$.
A musical element analysis data consisting of data for the structure and form of the song is extracted from a *Standard MIDI File (SMF)* as digitized music data. We form a musical element vector $\mathbf{s}$ which consists of music elements $key, tempo, pitch, rhythm, harmony$ and $melody$ generated from the musical element analysis data.

The vector is represented as follows.

$$\mathbf{s} = (key, tempo, pitch, rhythm, harmony, melody)^t. \quad (6)$$

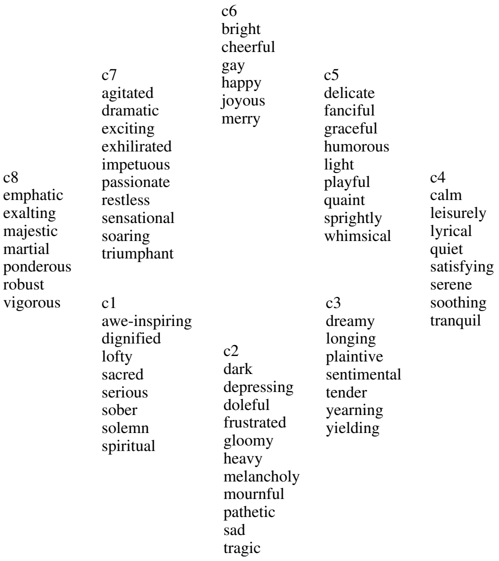

Figure 3: Hevner's 8 categories of impression words.

	key'	tempo'	pitch'	rhythm'	harmony'	melody'
$\mathbf{c}_1$	4	-14	-10	18	3	4
$\mathbf{c}_2$	-12	-12	-19	3	-7	0
$\mathbf{c}_3$	-20	-16	6	-9	4	0
$\mathbf{c}_4$	3	-20	8	-2	10	3
$\mathbf{c}_5$	21	6	16	8	12	-3
$\mathbf{c}_6$	24	20	6	-10	16	0
$\mathbf{c}_7$	0	21	-9	2	-14	-7
$\mathbf{c}_8$	0	6	-13	10	-8	-8

Figure 4: Transformation Matrix T indicating the relationships between impression word categories and musical elements.

Step 3 :Extraction of impression words
Transformation Matrix T transforms musical element vector $\mathbf{s}$ to the weights $\mathbf{v}$ (music category vector) of the 8 categories of impression words.

$$\mathbf{v} = T\mathbf{s} \quad (7)$$

A music category vector $\mathbf{v}$ is an 8 dimensional real valued vector.

$$\mathbf{v} = (v_{c_1}, v_{c_2}, \cdots, v_{c_8})^t. \quad (8)$$

The impression words in the same category are equally weighted by the corresponding weights of the category given by $\mathbf{v}$.

This is metadata due to weighted impression word categories, which is output by the media-lexicon transformation operator for music data. We can produce a set of weighted impression words from a music media data given in the form of MIDI.

3.3 Construction of Operator to Generate Facial Expression

A set of construction operators to generate facial expression for each fundamental feeling (basic emotion) such as happiness, surprise, fear, anger, disgust, sadness is based on Facial Action Coding

Table 1: A part of Action Unit.

AU	Description
1	Inner Brow Raiser
2	Outer Brow Raiser
4	Brow Lower
5	Upper Lid Raiser
6	Cheek Raiser
7	Lid Tighter
9	Nose Wrinkler
10	Upper Lip Raiser
12	Lip Corner Puller
15	Lip Corner Depressor
17	Chin Raiser
20	Lip stretcher
23	Lip Tighter
24	Lip Pressor
25	Lips part
26	Jaw Drop
27	Mouth Stretch

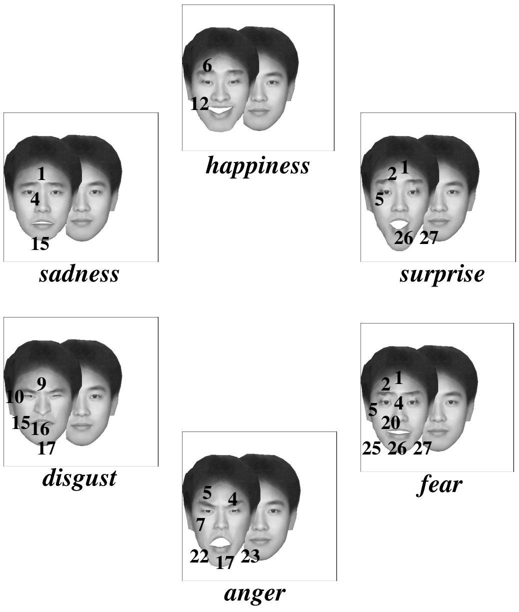

Figure 5: Relations between basic emotions and AU

System(FACS)(P.Ekman et.al 1978, Ekman & Friesen 1987).

3.3.1 Facial Action Coding System

Facial Action Coding System (FACS) by research of P. Ekman and W.V. Friesen can be used for showing a motion of a face based on dissection analysis of action of a face. In an objective facial expression consultation system, this is one of the methods currently used strictly and most widely.

P. Ekman and W.V. Friesen have shown how the appearance of a face changes with expansion and contraction of the each part of a face. And, they have clarified the method of determining how each emotion relates to each part of a face.

FACS describes facial expression with the combination of some Action Units (AU's). AU's are the minimum units of facial expression operations which are visually discernible. Table 1 shows a part of explanation of AU identifiers and their operations.

Moreover, the research results(P.Ekman et.al 1978, Ekman & Friesen 1987) have shown some combination of AU's for expressing each basic emotion. Figure 5 shows typical combination. The numbers in each face in Figure 5 express the identifiers of AU's, and the face in the back expresses the expressionless face with which feeling is not expressed. These facial expression are constructed using "Face Tool"(FaceTool).

The basic emotions are further arranged in a circle so that emotions are adjacent to other emotions to which they have similarities. The combination of these basic emotions can express a complicated facial expression.

3.3.2 An Implementation Method for Construction of Facial Expression

This section shows an implementation method for construction of facial expression. This operator consists of three steps:

Step 1 : Composition of Transformation Matrix F.
Transformation Matrix F is composed by the correlations between each basic emotion and Action Units (AU's) which are the minimum units of facial expression. The correlations are presented in (P.Ekman et.al 1978, Ekman & Friesen 1987) and shown in Figure 5.

Step 2 : Composition of a basic emotion vector $\mathbf{w}$.
We form a basic emotion vector $\mathbf{w}$ which has the weights of the basic emotion, and the vector is defined as

$$\mathbf{w} = (w_1, w_2, \cdots, w_6)^T. \qquad (9)$$

Step 3 :Construction of facial expression
Transformation Matrix F transforms a basic emotion vector $\mathbf{w}$ to the weights $\mathbf{u}$ of the AU's.

$$\mathbf{u} = F\mathbf{w} \qquad (10)$$

It is possible to construct an expression corresponding to each basic emotion by reflecting the weights $\mathbf{u}$. We can construct facial expression from basic emotions.

4 Visualization of Music Impression in Facial Expression to Represent Emotion

This section shows a visualization method of music impression in facial expression to represent emotion. This method can measure the relationship between each impression of music data and face expression. In section 4.1, we represent an associative heterogeneous mediadata search method. In section 4.2, we propose a visualization method of music impression in facial expression to represent emotion.

4.1 An Associative Heterogeneous Mediadata Search Method

An associative heterogeneous mediadata search method is shown in Figure 6.

The media-lexicon transformation operator is applied to extract impression words from each mediadata. The mathematical model of meaning(Kitagawa & Kiyoki 1993, Kiyoki, Kitagawa & Hayama 1994) is applied to extract semantically related each word. Therefore this model can measure the relation of each word extracted from media-lexicon transformation operator for each mediadata. By these functions,

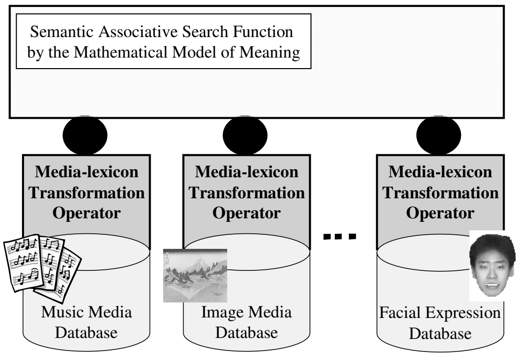
Semantic Associative Search Function
by the Mathematical Model of Meaning
Media-lexicon Transformation Operator
Media-lexicon Transformation Operator
Media-lexicon Transformation Operator
Music Media Database
Image Media Database
Facial Expression Database

Figure 6: A fundamental framework for an associative heterogeneous mediadata search method.

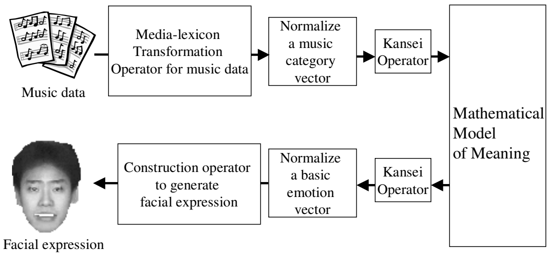
Music data
Media-lexicon Transformation Operator for music data
Normalize a music category vector
Kansei Operator
Mathematical Model of Meaning
Kansei Operator
Normalize a basic emotion vector
Construction operator to generate facial expression
Facial expression

Figure 7: The process of a visualization method of music impression in facial expression to represent emotion.

this method can measure the relationship between each impression of heterogeneous mediadata.

By realization of this method, an integration appropriately corresponding to the impression between heterogeneous mediadata on meta-level are realized easily. This method can generate new information by the integration appropriately corresponding to the impression between heterogeneous mediadata on meta-level. This method realizes to bridge over heterogeneous mediadata which exist independently as different database resources.

4.2 An Visualization Method of Music Impression in Facial Expression to Represent Emotion

In this section, we show a visualization method of music impression in facial expression to represent emotion. This method can composite facial expression corresponding to impressions of a music data.

Facial expression is one of nonverbal behaviors. The facial expression is important as media which convey various emotion effectively. Assuming that an integration between existing mediadata and facial expression is realized, the interface corresponding to human Kansei is realized.

The process of this method is shown Figure 7.

This method consists of following operations.

1. Mathematical model of meaning
   The Mathematical model of meaning can measure the semantic relation between each word. This function measures correlations between 6 basic emotion words and 8 impression word categories with weights from the media-lexicon transformation operator for music data.

   This model has shown in section 2.

2. Media-lexicon transformation operator for music data
   The media-lexicon transformation operator for music data can extract the weighted impression word categories corresponding to the impression of the music data.

   This function has shown in section 3.2.

3. Normalize a music category vector
   Impression word categories corresponding to the music data and their weights are output by the media-lexicon transformation operator. However, these weights generally have not been normalized.

   The following formulas f_N are applied in this paper.

   $$f_N(v_{c_1}, v_{c_2}, \cdots, v_{c_8}) : (v'_{c_1}, v'_{c_2}, \cdots, v'_{c_8}) \mapsto (\frac{v_{c_1}}{max_1}, \frac{v_{c_2}}{max_2}, \cdots, \frac{v_{c_8}}{max_8}). \quad (11)$$

   $max_1, max_2, \cdots, max_8$ denote the maximum weight values of each words category. Details are presented in reference (Kitagawa, Nakanishi & Kiyoki 2004).

4. Construction operator to generate facial expression
   The construction operator to generate facial expression can construct facial expression from the basic emotion vector which is correlation values measured in the mathematical model of meaning.

   This function has shown in section 3.3.

5. Normalize a basic emotion vector
   The basic emotion vector which consists of 6 correlations of basic emotions for construction of facial expression is extracted by measurement of the relation between basic emotion words and impression words of music data in the mathematical model of a meaning. However, the basic emotion vector generally has not been normalized. The following formulas are applied in this paper.

   The basic emotion vector **fev** which consists of 6 non-normalization correlations extracted by the mathematical model of meaning is defined as

   $$\mathbf{fev} = (b_1, b_2, \cdots, b_6)^T. \quad (12)$$

   This vector is normalized as follows:

   $$\mathbf{fev}' = (b'_1, b'_2, \cdots, b'_6)^T. \quad (13)$$
   $$b'_i = \frac{b_i}{\sum_{j=1}^{6} b_j}.$$

   It is shown in reference (Ekman & Friesen 1987) that a complicated facial expression can be constructed with the combination of 6 basic emotions. Each AU is independent anatomically. These formulas are normalization to the value showing the rate which each basic emotion combine.

   Moreover, there is a risk that the small amount of features may generally be impurities which worsen results. The feature of facial expression is more clarified by removing these values. These are shown as follows, using the removed value as w_i.

$$w_i = \begin{cases} b'_i & (b'_i \geq \varepsilon) \\ 0 & (b'_i < \varepsilon) \end{cases}$$

$$\varepsilon = \frac{\sum_{j=1}^{6} b'_j}{6}. \quad (14)$$

Thus the normalized basic emotion vector is constituted.

$$\mathbf{w} = (w_1, w_2, \cdots, w_6)^T. \quad (15)$$

These formulas are not always determined in this normalization method, because there is a limit in finding suitable normalization formulas in an experiment. These formulas as the normalization method are open to discussion. These normalization formulas need verification by the specialist. This verification is a future work.

6. Kansei operator

   The Kansei operator(Kitagawa, Nakanishi & Kiyoki 2004) is used to adjust the expression with the interpretation of human sensitivity computed by the logarithmic function based on Fechner's law(Ohyama et al. ed. 1994). This function and a semantic associative search method make it possible to realize semantic search according to the human Kansei for multimedia data.

   When impression word weights are composed for each feature, Kansei operator positions the sum total of each of the features as the stimulus strength and uses Fechner's law to obtain the sensation magnitude corresponding to that stimulus as the composed weight.

   (a) Fechner's law

   E.H. Weber has shown that human beings perceive the ratio of the difference in the magnitudes of objects rather than perceives the difference between the magnitudes of objects by the discrimination experiment of weights. Fechner named this fact, which Weber had discovered, Weber's law.

   Fechner supposed that Weber's law is generally applied and leads to

$$d\gamma = k\frac{d\beta}{\beta}, \quad (16)$$

   where k : proportionality constant; β : magnitude of a stimulus(hereafter stimulus strength); γ : sensation magnitude; $d\beta, d\gamma$: infinitesimal increases in the stimulus strength and sensation magnitude.

   By the integration of (16),

$$\gamma = k(\log\beta - \log b), \quad (17)$$

   where $\log b$ is the integration constant. Therefore, γ is as follows:

$$\gamma = k\log\frac{\beta}{b}. \quad (18)$$

   The sensation magnitude is proportional to the logarithm of the stimulus strength. This is called the Fechner's law.

   (b) Construction of the Kansei operator

   Each feature assigned to each impression word can be viewed as a stimulus in that feature. Obtaining the sum totals of each feature can be thought of as obtaining the stimulus strength in each feature possessed by the mediadata Therefore, the sum total of each feature in each impression word can be assigned the meaning of the stimulus strength of that feature.

   Kansei operator g is shown as follows;

$$\mathbf{y} = (y_1, y_2, \cdots, y_n)^T,$$
$$g(\mathbf{y}) := (\gamma_1, \gamma_2, \cdots, \gamma_n)^T, and$$
$$\gamma_j = \begin{cases} k\log_\alpha|y_j| + 1 & (y_j > 0) \\ 0 & (y_j = 0) \\ -(k\log_\alpha|y_j| + 1) & (y_j < 0) \end{cases} \quad (19)$$

   where k and α are the parameters which can be set up as sensational volumes. These parameters are taken as $k = 1$, $\alpha = 6$ in the case of music data (Kitagawa, Nakanishi & Kiyoki 2004), and $k = 10$, $\alpha = 18$ in the case of facial expression by our pilot studies.

5 Experiments

To verify the effectiveness of this method, we built an experimental system based on this method, and performed verification experiments.

5.1 Experimental environment

To create metadata space $\mathcal{MDS}$, we used the English-English dictionary *Longman Dictionary of Contemporary English*(Sumners et al. ed. 1987). This dictionary uses only approximately 2,000 basic words to explain approximately 56,000 headwords. We created the data matrix M in subsection 1 of section 2 by treating basic words as features and setting the element corresponding to a basic word to "1" when the basic word explaining a headword had been used for an affirmative meaning, setting it to "-1" when the basic word had been used for a negative meaning, setting it to "0" when the basic word was not used, and setting it to "1" when the headword itself was a basic word. In this way, we generated the metadata space $\mathcal{MDS}$, which is an orthonormal space of approximately 2000 dimensions. This space can express 2^{2000} different phases of the meaning.

The facial expression construction part in this experiment system is realized by utilizing "Face Tool"(FaceTool).

5.2 Experimental Method

We verify output results which are the generated facial expression by the some music data in this experimental system.

We use 4 well-known pieces which have rather apparent impression to anyone in MIDI format for some input data in this system. These pieces are "Clap your hands", "Brahms 3rd Symphony", "Song of four seasons", and "Silent night holy night". The well-approved impressions of the pieces are as follows: "Clap your hands", which goes like clap your hands if you are happy, has impression of merry, happy, and joy. "Brahms 3rd Symphony" has impression of heavy. "Song of four seasons", which goes like one who loves spring has pure heart, has impression of tender, sad and sentimental. "Silent night holy

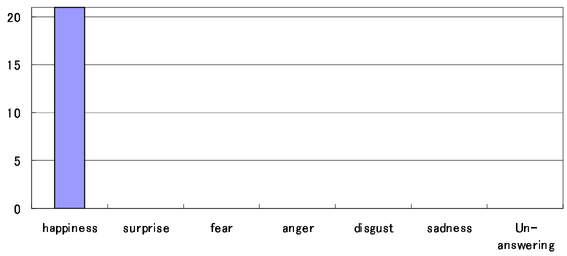

Figure 8: The result of subject investigation about impression of “Clap your hands”.

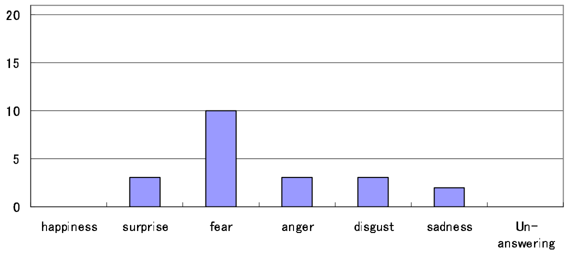

Figure 9: The result of subject investigation about impression of “Brahms 3rd Symphony”.

night” has impression of holy and solemn. This experiment system inputs these music data, and 4 facial expressions corresponding to those impressions are extracted in this experiment. We conduct hearing investigations about impressions of these facial expressions and impressions of these music data.

5.3 Experiment Results

First, hearing investigation was conducted about impressions of these music data. Subjects of 21 adult men and women are asked to select the appropriate impression corresponding to music data in 6 items. The items are “happiness”, “surprise”, “fear”, “anger”, “disgust”, and “sadness”. These items are the same as basic emotions for facial expression.

The results of these investigations about each impression about “Clap your hands”, “Brahms 3rd Symphony”, “Song of four seasons”, and “Silent night holy night” are shown in Figure 8, 9, 10, and 11.

In case of “Clap your hands” shown in Figure 8, all subjects have answered “happiness”. This result corresponds to the impression of this music that we assumed.

In case of “Brahms 3rd Symphony” shown in Figure 9, more than 40% subjects have answered “fear”. However other subjects have answered “surprise”, “anger”, “disgust”, and “sadness”. We assumed the impression of this music to be heavy. The impression of “heavy” is close on sadness, angry, and fear in meaning. Thus this result corresponds to the impression that we assumed.

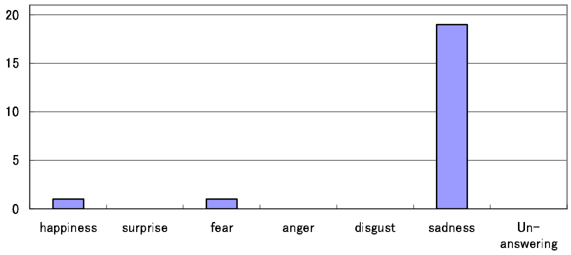

Figure 10: The result of subject investigation about impression of “Song of four seasons”.

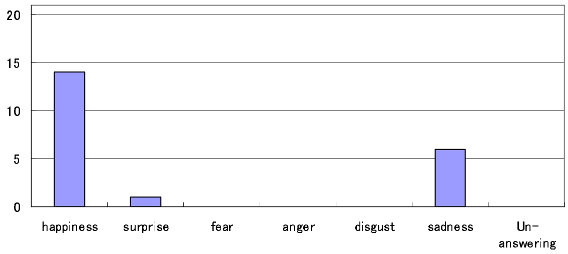

Figure 11: The result of subject investigation about impression of “Silent night holy night”.

Table 2: The result extracted from “Clap your hands”.

C6	59.019306
C5	37.261279
C7	13.467024
C8	-6.346211
C4	-11.518259
C3	-17.185234
C1	-25.459551
C2	-34.546004

In case of “ Song of four seasons” shown in Figure 10, more than 90% subjects have answered “sadness”. This result corresponds to the impression of this music that we assumed.

In case of “Silent night holy night” shown in Figure 11, more than 60% subjects have answered “happiness”, and about 30% subjects have answered “sadness”. This result never corresponds to the impression of this music that we assumed.

The results extracted by the media-lexicon transformation operator for music shown in section 3.2 from 4 MIDI data are shown in the Table 2, 3, 4, and 5.

In case of “Clap your hands” shown in Table 2, weight of the impression word category of C6 expressing “happy” is the largest. This result corresponds to the subject investigation.

In case of “Brahms 3rd Symphony” shown in Table 3, weights of the impression word categories of C2 and C3 expressing “sad”, “heavy”, and “longing” are large. The impression of “heavy” is close on sadness, angry, and fear in meaning. This result almost corresponds to the subject investigation.

In case of “ Song of four seasons” shown in Table 4, weights of the impression word categories of C2 and C3 expressing “sad”, “heavy”, and “longing” are large. This result corresponds to the subject investigation.

In case of “Silent night holy night” shown in Table 5, weight of the impression word category of C1 expressing “serious” is the largest. However, weight of the impression word category of C6 expressing “happy” is also large. This song has not only serious but also happy. The impression word category of C2

Table 3: The result extracted from “ Brahms 3rd Symphony ”.

C2	21.727009
C3	14.127015
C7	6.552777
C8	-4.168117
C4	-8.970285
C6	-18.778780
C5	-19.791271
C1	-19.205445

Table 4: The result extracted from “ Song of four seasons”.

C3	17.863154
C2	14.470195
C4	1.070744
C7	-2.208193
C8	-7.584915
C5	-9.435428
C6	-9.925302
C1	-17.619686

Table 5: The result extracted from “ Silent night holy night”.

C1	21.469676
C6	11.058502
C5	9.959276
C4	6.508946
C8	5.181874
C7	-6.144448
C2	-9.816415
C3	-12.011969

expressing “sadness” which about 30% subjects have answered has negative weights. This result never corresponds to the subject investigation. We find that it is difficult for such song to decide impression.

The results of measurement of correlations between 6 basic emotion words and 8 impression word categories with weights utilizing the mathematical model of meaning are shown in the Table 6, 7, 8, and 9.

In case of “Clap your hands” shown in Table 6, correlation of “happiness” is the largest in 6 basic emotion words. “Happiness” is close to “C6” semantically. In case of “Brahms 3rd Symphony” shown in Table 7, correlation of “sadness” is the largest in 6 basic emotion words. “Sadness” is close to “C2” or “C3” semantically. In case of “ Song of four seasons” shown in Table 8, correlation of “sadness” is also the largest in 6 basic emotion words. In case of “Silent night holy night” shown in Table 9, correlations of “happiness”, “surprise” and “anger” are large in 6 basic emotion words. “C1” is close to “anger” and “surprise” in the mathematical model of meaning utilizing the space constructed by Longman Dictionary of Contemporary English(Sumners et al. ed. 1987). “C1” is closer to “C8” expressing “emphatic” than “C4” expressing “calm” in Hevner’s work. Actually, “anger” and “surprise” are emphatic expressions. These results are appropriately measured by the mathematical model of meaning.

Finally, the experimetal results which facial expression are constructed are shown in Figure 12, 13, 14, and 15.

In the case of Figure 12, “happiness” as facial expressions are automatically created. In the case of Figure 13, facial expression which mixed “sadness” and “fear” is automatically created. In the case of Figure 14, “sadness” as facial expressions are automatically created. In the case of Figure 15, facial expression which mixed “happiness” and “surprise” are automatically created.

Table 6: The result of measurement correlations (“Clap your hands”).

happiness	0.153885
anger	0.134517
surprise	0.131953
fear	0.127711
disgust	0.105515
sadness	0.077445

Table 7: The result of measurement correlations (“Brahms 3rd Symphony”).

sadness	0.340237
anger	0.177366
fear	0.173911
surprise	0.143764
disgust	0.123588
happiness	0.097354

Table 8: The result of measurement correlations (“Song of four seasons”).

sadness	0.326680
anger	0.214740
fear	0.204385
disgust	0.190738
surprise	0.183363
happiness	0.139167

Moreover, the results by subjects of 21 adult men and women are shown Figure 16, 17, 18, and 19. These results show whether the output result corresponds to the impression of input pieces. All subjects are asked to select the most appropriate item from “Exaggerated”, “Correct”, “Almost Correct”, “Comfortable”, “Slightly Different”, and “Totally Different”.

In the case of “Clap your hands” shown in Figure 16, the more than 90% subjects have answered “Correct”, “Almost Correct” and “Comfortable.” This result is shown that this facial expression corresponds to impressions of this song.

In the case of “Brahms 3rd Symphony” shown in Figure 17, the more than 70% subjects have answered “Correct”, “Almost Correct” and “Comfortable.” This result is shown that this facial expression corresponds to impressions of this song.

In the case of “Song of four seasons” shown in Figure 18, the more than 80% subjects have answered “Correct”, “Almost Correct” and “Comfortable.” This result is shown that this facial expression corresponds to impressions of this song.

In the case of “Silent night holy night” shown in Figure 18, the more than 70% subjects have answered “Exaggerated”, “Slightly Different” and “Totally Different”.

The media-lexicon transformation operator for music extracts not only “C1” expressing “serious” but also “C6” expressing “happy” from “ Silent night holy night”. Hereby, facial expression which mixed “happiness” and “surprise” are automatically created. These results are appropriate for the experimental system. In contrast, “Silent night holy night” has the more than 60% subjects who have answered “happiness” like “Clap your hands”. However, the impression of “Silent night holy night” is holy and solemn unlike “Clap your hands” certainly. In the case of this song, it is thought that the impression of other elements such as lyrics, a title and so on

Table 9: The result of measurement correlations (“Silent night holy night”).

happiness	0.228688
surprise	0.185209
disgust	0.178705
anger	0.176306
fear	0.175334
sadness	0.114702

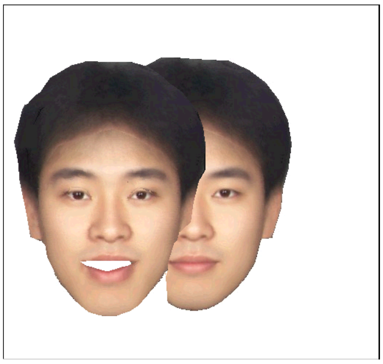

Figure 12: The result ("Clap your hands").

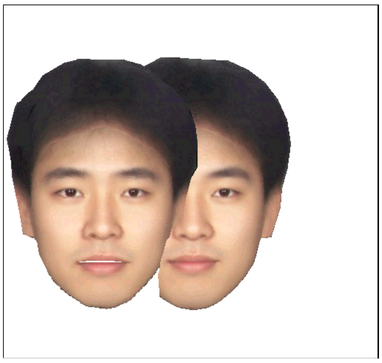

Figure 13: The result ("Brahms 3rd Symphony").

is large. Assuming that a media-lexicon transformation operator for other elements is realized, clearer impression words will be extracted and appropriate facial expression will be constructed. A realization of the media-lexicon transformation operator for other elements is our future work.

As shown in those results, we have clarified that our method constructs various facial expressions from arbitrary music data.

6 Conclusion

In this paper, we proposed a visualization method of music impression in facial expression to represent emotion. We clarified the effectiveness of this method by showing several experiment results.

This method realizes an integration corresponding to the impression between existing mediadata and nonverbal behaviors that convey various emotions effectively. The interface corresponding to human Kansei with less difficulty for a user is realized by this method which realize a integration appropriately corresponding to the impression between existing mediadata and facial expression.

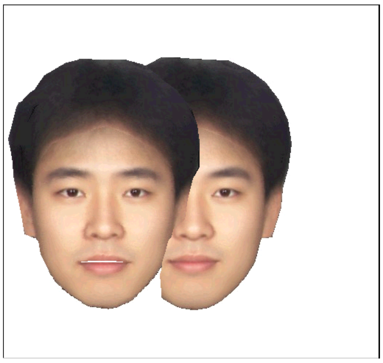

Figure 14: The result ("Song of four seasons").

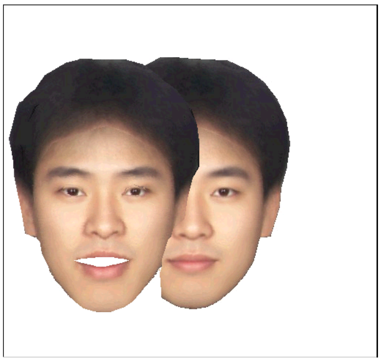

Figure 15: The result ("Silent night holy night").

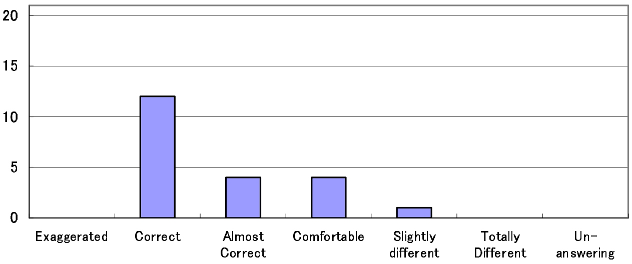

Figure 16: The result of subject investigation about "Clap your hands".

We believe that this method which realizes an integration appropriately corresponding to the impression between existing mediadata and facial expression can be used for a realization of the interface corresponding to human Kansei with less difficulty for a user.

As our future work, we will realize a learning mechanism according to individual variation. We will also consider analytical evaluation and verification by the specialist for facial studies. Furthermore, we will apply this method to various type of existing mediadata and nonverbal behaviors.

Acknowledgment

We really appreciate Prof. Yasushi Kiyoki, Faculty of Environmental Information, Keio University, for his substantial supports to enhance the presentation of the paper.

References

Ekman,P. & Friesen,W.V. (1978), Facial Action Coding System, *Consulting Psychologist Press.*

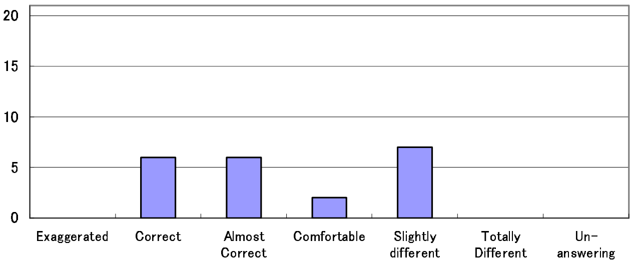

Figure 17: The result of subject investigation about "Brahms 3rd Symphony".

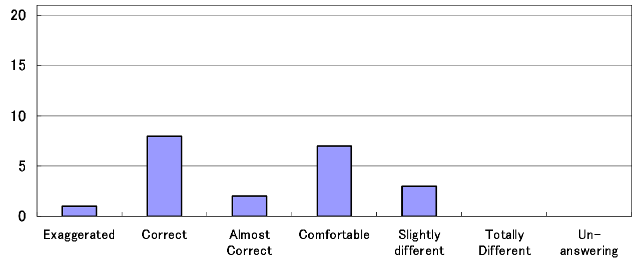

Figure 18: The result of subject investigation about "Song of four seasons".

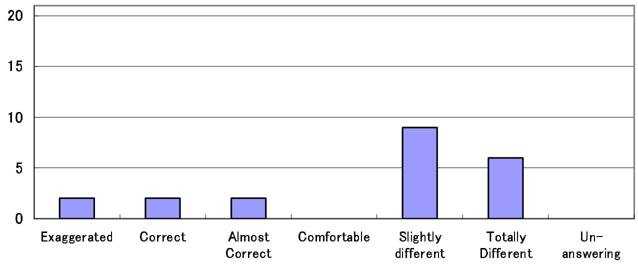

Figure 19: The result of subject investigation about "Silent night holy night".

Ekman,P. & Friesen,W.V., Kudo,T.(translator and editor) (1987), A guide to expression analysis-The meaning hidden in expression is explored, *Sisin-Shobou Press.*

Choi,CS., Harashima,H. & Takebe,T. (1990), 3-dimensional facial model-based description and. synthesis of facial expressions, *The Transactions of the Institute of Electronics, Information and Communication Engineers*, Vol.J73-A, No.7, pp.1270-1280.

Hara,F. & Kobayashi,H. (1996), Real-time facial interaction between human and 3D face agent, *Proc.5th IEEE International Workshop on Robot and Human Communication (RO-MAN'96)*, pp.401–409.

Kitagawa,T. & Kiyoki,Y. (1993), The mathematical model of meaning and its application to multidatabase systems, *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp.130–135.

Kiyoki,Y., Kitagawa,T. & Hayama,T. (1994), A metadatabase system for semantic image search by a mathematical model of meaning, *ACM SIGMOD Record*, vol. 23, no. 4, pp.34–41.

Kitagawa,T. & Kiyoki,Y. (2001), Fundamental framework for media data retrieval system using media lexico transformation operator, *Information Modelling and Knowledge Bases*, vol.12, pp. 316–326.

Kitagawa,T., Nakanishi,T. & Kiyoki,Y. (2004), An Implemantation Method of Automatic Metadata Extraction Method for Music Data and its Application to a Semantic Associative Search, *Systems and Conputers in Japan*, Vol.35, No.6, pp59-78.

Hevner,K. (1935), Expression in music: A discussion of experimental studies and theories, *Psychological Review*, Vol. 42, pp. 186–204.

Hevner,K. (1936), Experimental studies of the elements of expression im music, *American Journal of Psychology*, Vol. 48, pp. 246–268.

Hevner,K. (1937), 'The affective value of pitch and tempo in music, *American Journal of Psychology*, Vol. 49, pp. 621–630.

Umemoto,T.(editor). (1966), Music Psychology, *Seishin-Shobo Press.*

FaceTool http://www.hc.t.u-tokyo.ac.jp/project/face/

Oyama,T., Imai,S. & Wake,T.(editors) (1994), New edition, Handbook of sensory and perceptive psychology, *Seishinshobou Press.*

Sumners,D. et al. (1987), Longman dictionary of contemporary English, *longman.*