

A Kansei Model for Musical Timbre Using a Neural Network

Gen ONISHI

Department of Mechanical and Electrical Engineering, Tokuyama College of Technology
3538 Kumetakajyou Tokuyama-shi, Yamaguchi 745-8585, Japan

Masatoshi NIIZEKI

Department of Electro-Mechanical Engineering, Osaka Electro-Communication University
18-8 Hatsu-cho, Neyagawa-shi, Osaka 572-8530, Japan

Ichiro KIMURA

Department of Electro-Mechanical Engineering, Osaka Electro-Communication University

Hidemi YAMADA

Department of Mechanical and Electrical Engineering, Tokuyama College of Technology

ABSTRACT

This paper presents a "Kansei" model for musical timbre using a neural network. The three-layered network, which is constructed based on the structure of auditory system, consists of seven input nodes, four neurons in the hidden layer, and three neurons as the output nodes. The input is a pattern of tone pressure levels for harmonics and the output is a three-dimensional vector of the Kansei value that characterizes "clearness-unclearness," "thickness-thinness" and "sharpness-softness." After training the network, the analysis of the synaptic weights of all the neurons gives the following results that agree well with knowledge obtained from the field of psychology and physiology. (1) Each harmonic equally influences the Kansei value for "thickness-thinness." (2) Harmonic components which are higher than the fourth harmonic have an inverse effect on the Kansei value for "clearness-unclearness." (3) Two neurons that have a lateral inhibition property exist in the hidden layer, and these neurons extract the sixth and seventh harmonics. Another neuron in the hidden layer extracts the fifth, sixth and seventh harmonics. These three neurons determine the Kansei value for "sharpness-softness."

Keywords: Neural Network, Kansei, Musical Timbre, Harmonic, Hearing mechanism, modeling

1. INTRODUCTION

The products that we encounter in our daily life these days must have high quality and must be high-performance. We now prefer products that feel comfortable and are fitting to our tastes [1][2][3]. A human Kansei model receives and stores the preferences that people have toward certain phenomena to generate the many different engineering parameters necessary in the design of these products. There has been

considerable research on statistical methods for constructing a computer database to establish design factors such as shapes, colors and patterns from human Kansei [4][5]. The Kansei of a human being responds to musical stimuli in many ways. This response results in a person having various feelings such as being happy or sad. There have been many successful studies concerning the reaction of human Kansei for musical input. In a previous report, the authors constructed a Kansei model based on the human auditory system using a neural network to obtain Kansei values from musical chords. Furthermore, we confirmed that the model has successfully simulated human Kansei and its characteristics agree well with knowledge obtained from the fields of psychology and physiology [6]. The final purpose of this study is to construct a human Kansei model that simulates the human response to the input of a set of comprehensive musical values which represent all of the possible types of musical elements. The Kansei of a human being responds to musical timbre, one of musical elements. The improvement of the musical timbre for instruments and audio products will lead to a more comfortable sound environment.

In this study, we construct a Kansei model for musical timbre using a neural network. The input of the network is musical timbre, and the output is a set of Kansei values which characterizes "clearness-unclearness," "thickness-thinness" and "sharpness-softness." An analysis of the synaptic weights of the neurons after the training of the network shows the validity of this model.

2. THE DEFINITION OF MUSICAL TIMBRE

Musical timbre is determined by its harmonic structure. The IEEE dictionary defines timbre as follows:

"The attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar. Timbre depends primarily upon the spectrum of the

stimulus, but it also depends upon the waveform, the sound pressure, the frequency location of the spectrum, and the temporal characteristics of the stimulus [7]."

In this definition, the loudness and the fundamental frequency (pitch) for each timbre must be kept constant, when we distinguish the difference of the two timbres. But we can perceive a guitar sound as a guitar sound, and we can recognize the difference between guitar sounds and flute sounds, whether the loudness and the pitch for each timbre is kept constant or not [8].

The timbre of each musical instrument has a different harmonic structure, a different time series of harmonic structures and a different time series of sound pressure levels. These factors enable us to recognize the difference in the sounds of the musical instrument. If we construct timbre values by considering these factors, we would need a huge number of timbre values. For example, if we play the guitar two times with the same pitch and the same loudness, the timbre is not exactly the same. This is due to slight differences in the time series of the harmonic structure and sound pressure level resulting from variations in attacking time and attacking intensity of the guitar string. In spite of this, we perceive that the timbre is almost the same.

Two alternative definitions of musical timbre are given in [8].

- (1) Musical timbre in the narrow sense, is the property of sound which we perceive from its harmonic structure.
- (2) Musical timbre in the broad sense, is the perception which we obtain from the loudness and pitch of the sound that are caused by the various amplitudes and dynamic or spatial combination of pure tone frequencies.

In this study, we define musical timbre in the narrow sense because of its simplicity, and because it embraces all of the most basic qualities of sound.

3. ACQUISITION OF MUSICAL VALUES AND KANSEI VALUES

Construction of Musical Values

Musical timbre is constructed by the combination of sine waves which each have different frequencies. Figure 1 shows the waveform and the spectra of a triangle wave, and its timbre is similar to the musical timbre of string and brass instruments. This figure shows that a timbre is consisted by the fundamental tone and the sine-waves of the harmonic tone frequencies. The fundamental tone of the sound is the sine-wave with the lowest frequency. A sound with a distinct pitch such as the sound of a musical instrument consists of a fundamental frequency and its integer multiples. There is a perceivable pitch in the timbre of the sound [10]. On the other hand, a sound in nature contains many non-integer multiple frequencies of the fundamental tone. The sound with many non-integer multiple frequencies has no distinct musical pitch, and is perceived to be noise. Since it is not possible to represent all of the vast number

of sounds in nature, this study limits the timbre samples to musical tones. Though non-integer harmonics are also included in musical tones, the proportion is much smaller than the integer harmonic. Furthermore, the sound pressure levels of the non-integer harmonics are very small. In fact, human beings do not perceive the non-integer harmonics of the musical tones. At least six of the harmonics are necessary to distinguish between each musical timbre [10]. Human beings are able to perceive five to seven of the harmonics in a musical tone [11]. For reasons mentioned above, musical tones in this study are steady compound tones which consist of the fundamental frequency and the harmonics from the second to the seventh.

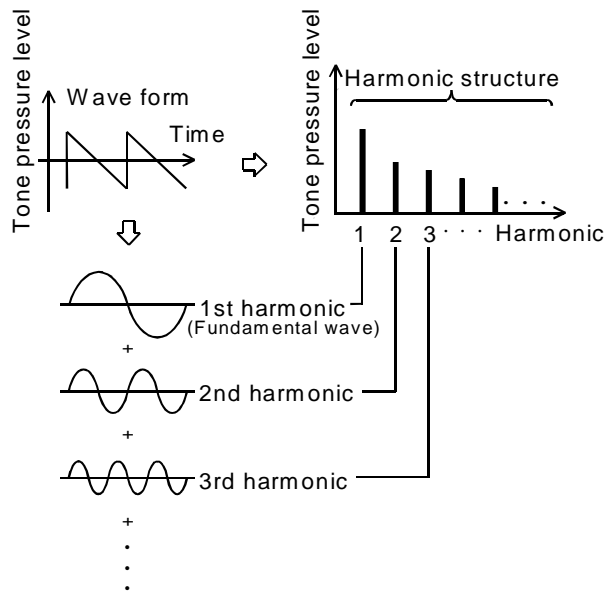


Fig.1 Harmonic structure [9]

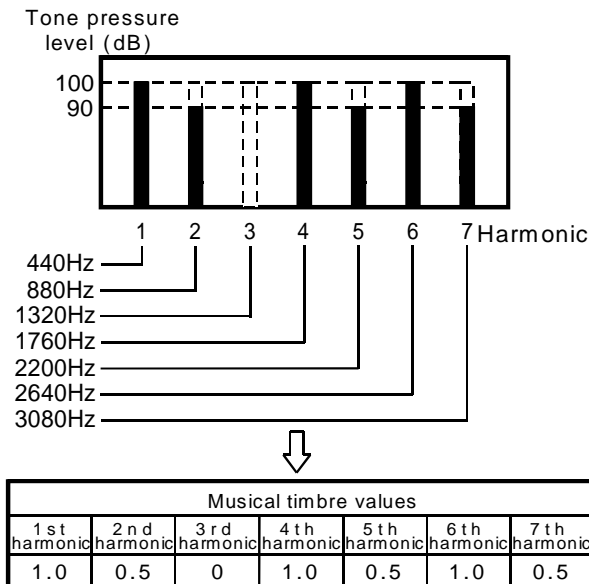


Fig.2 An example of musical timbre values

Figure 2 shows an example of a musical timbre. As described in chapter 2, difference in sound pitch has little influence on the perception of musical timbre. Therefore, the fundamental frequency of each musical timbre is set constant to 440 Hz. Furthermore, the range of many musical instruments covers this frequency and it is used as a standard for the tuning of many musical instruments [12]. If musical timbres have harmonics with different phases, then they have different waveforms. But even if the waveforms are different, humans perceive these to have the same musical timbre [10].

Therefore we use a pure tone (a sine wave consisting of a single frequency component) as the fundamental tone and the harmonic tones. The harmonics are set to be in-phase with the fundamental tone. The sound pressure level of each harmonic is adjusted by measuring the output sound of a speaker with a sound level meter. To equalize the musical values, the playing time of a musical timbre is set at two seconds, and the pressure level of the fundamental tone is set to either 90 or 100dB. The other harmonics are either turned off, or set to have a tone pressure level of 90 or 100dB, using a MIDI software. These tone pressure levels were selected since humans perceive sounds of 90dB to have about half the loudness of a sound of 100dB [10][13]. The input values for sound pressure when the sound is off, 90dB and 100dB, are set to 0.0, 0.5 and 1.0, respectively. The sound pressure level of each harmonic is combined to create 256 arbitrarily selected timbre data for input to the neural network as musical values.

Acquisition of Kansei values

The Kansei values that are obtained from a subjective test are the training values of the neural network. Five examinees listen to the 256 musical timbre patterns from a MIDI tone generator (Roland SK-50) through two speakers in the subjective test. They evaluate each musical timbre using the three word pairs "clearness-unclearness," "thickness-thinness" and "sharpness-softness" on a five step rating scale. There have been many studies concerning evaluation words for musical timbre, and these studies show that these words can be categorized into three groups [14]:

- (1) The pleasant word group, which represents the beauty and the pleasure of each sound
- (2) The power word group, which represents the loudness and the thickness of each sound.
- (3) The metallic word group, which represents the shrill effect of each sound.

In this study, we chose words from each of these groups to provide a comprehensive evaluation for musical timbres. Five examinees listen to all of the musical timbre in advance to get acquainted to the underlying features, and then take the subjective test two times. The Kansei values for the five examinees obtained from the subjective test were standardized from -0.9 to 0.9 to be input as training data.

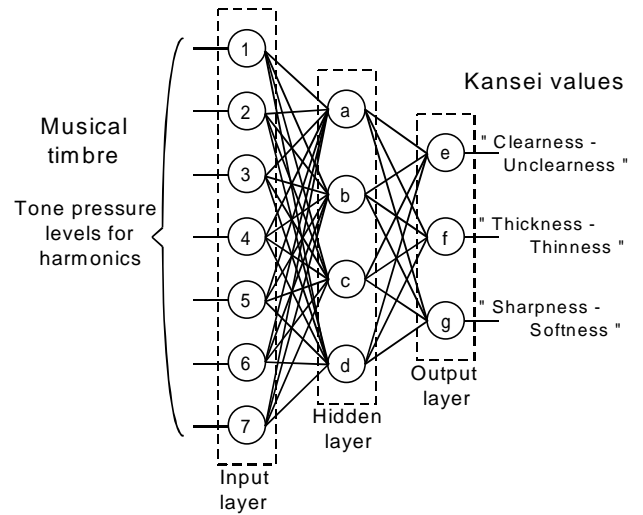


Fig.3 Network structure

4. THE STRUCTURE OF THE NETWORK MODEL

The network model takes a musical value as input and generates a Kansei value as output. Figure 3 shows the network model for this study. The input of the network model is the tone pressure levels (0.0 minimum to 1.0 maximum) of each harmonic in the musical timbre. The seven units in the input layer are set to the tone pressure of seven harmonic tones. The unit number represents the order of the harmonic. A Kansei value for "clearness-unclearness," "thickness-thinness" and "sharpness-softness" in the range (-0.9, 0.9) is obtained from output units *e*, *f* and *g* respectively. As a result of some tests, we determined that this network model functions best when it has a hidden layer consisting of four units. Equation (1) shows the sigmoid function used in this study, where *x* is the input of the neural network unit, and *f(x)* is the output. In the above manner, we have constructed a neural network model that takes musical timbre data as input and produces Kansei value as output in the range (-0.9, 0.9) for the evaluation words.

$$f(x) = \frac{2}{1 + e^{-x}} - 1 \quad (1)$$

5. NETWORK TRAINING AND PERFORMANCE EVALUATION

We train the network model with the back propagation algorithm, then the results of the training and the testing of the network model are evaluated by the following equation.

$$E = \frac{1}{2} \sum_{i=1}^n (y_i - t_i)^2 \quad (2)$$

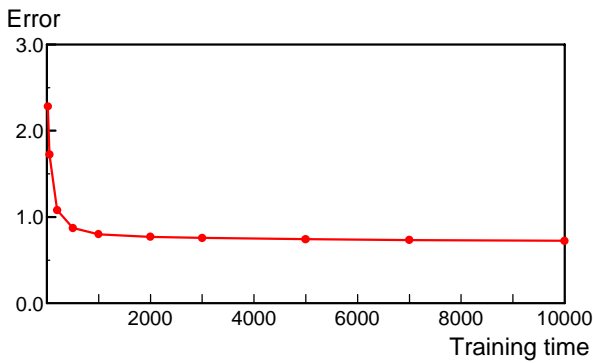


Fig.4 Training error curve

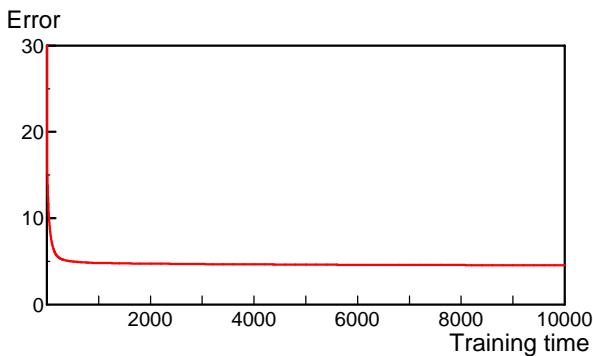


Fig.5 Testing error curve

E is the error function, y_i is the output of the network, t_i is the teaching values, n is the number of teaching values.

We arbitrarily choose 220 of the 256 musical timbres for training of the network, and the remaining 36 for tests.

Figure 4 shows the output error for the number of training times of the network. The output error of the network has smoothly settled without fluctuation so the network training is successful. Figure 5 shows the output error for the generalization test when the network training was terminated at a certain epoch. The errors of the training and the testing for the network converge within 10000 times. Therefore, we fixed the number of training sessions to 10000.

The results after training the network are shown in Figure 6. These graphs show the output Kansei values and the teaching Kansei values for each musical timbre. The Kansei values in graph (a), (b), (c) are "clearness-unclearness," "thickness-thinness" and "sharpness-softness," respectively. The results shown in Figure 6 confirm that the output Kansei values from the network well matches the values used in the training, and therefore the training is successful.

The results using the test data are shown in Figure 7. The results show that the output values also well matches the teaching Kansei values. Therefore, the network model has successfully simulated human Kansei for musical timbre.

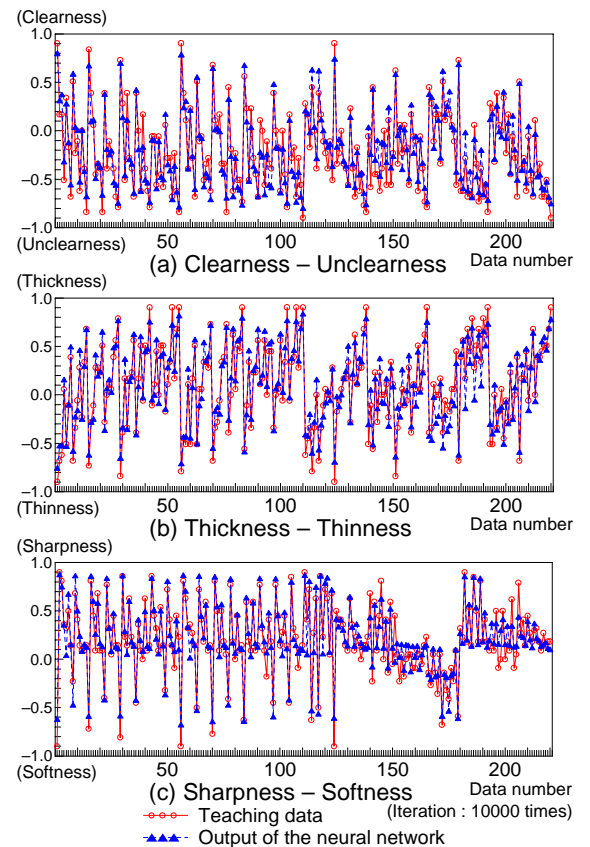


Fig.6 The results after training the network

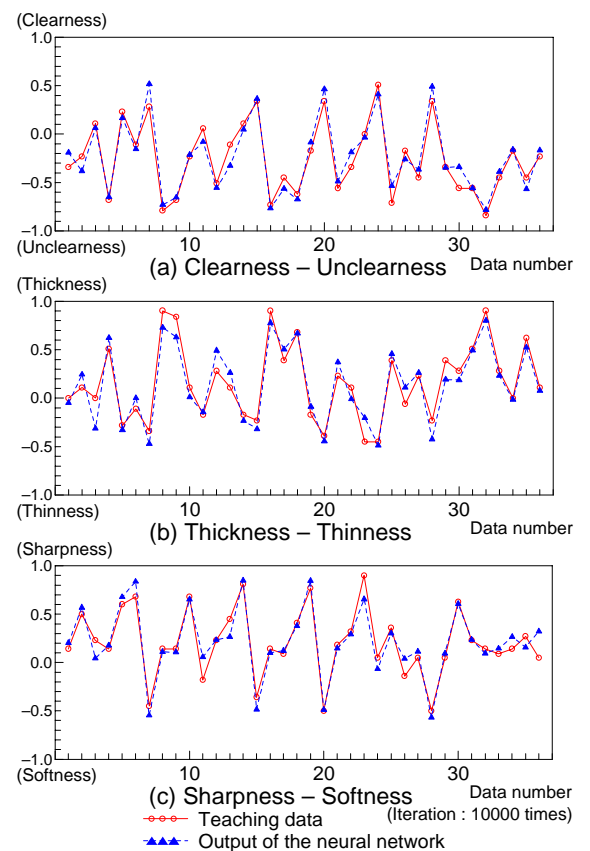


Fig.7 Test results

6. ANALYSIS OF THE SYNAPTIC WEIGHTS OF THE NETWORK MODEL

Analysis of the synaptic weights of the network after training reveals the units which most affect the output of each Kansei value. Figure 8 shows the synaptic weights and the bias of each unit in the network model after training. The white circles indicate positive synaptic weights and the black circles indicate negative synaptic weights. The sizes of the synaptic weights in figure 8 show their values.

Since the synaptic weight values from units *a* and *d* to output unit *e* are larger than other synaptic weights, the output of the Kansei values for "clearness-unclearness" is mainly determined by units *a* and *d*. Similarly, the "thickness-thinness" value is mainly determined by unit *a*, and the "sharpness-softness" value is mainly determined by units *b*, *c*, and *d*.

Next, we observe the synaptic weights between input layer units and hidden layer units. Every synaptic weight between units *a* and units 1, 2, 3, 4, 5, 6 and 7 are almost the same size, that is, unit *a* is equally supplied with the sound pressure levels from the fundamental tone to the seventh harmonic. Since the role of the negative synaptic weights for units *b* and *c* show a lateral inhibition property, unit *b* extracts the sixth harmonic (2640 Hz) and unit *c* extracts the seventh harmonic (3080 Hz). A particular neuron group in the inferior colliculus of the human auditory nervous system analyzes the difference between the frequencies of a sound with lateral inhibition [15][16][17]. Units *b* and *c* have obtained a lateral inhibition property, and these have a role similar to a band-pass filter.

In unit *d*, the synaptic weight values from units 5, 6 and 7 are larger than other synaptic weights. The results of the analysis for the network model indicate the following about each Kansei value. The Kansei value for "thickness-thinness" is determined by unit *a*, and all of the harmonics in a musical timber equally influence this Kansei value. A timber containing many harmonics is perceived by human beings as a rich sound [17][18].

The Kansei value for "clearness-unclearness" is determined by units *a* and *d*, and this Kansei value is acquired from the sum of the output of the two hidden units. Unit *a* is equally supplied with all harmonics in a sound, unit *d* mainly decides the Kansei value for "thickness-thinness." Harmonic components which are higher than the fourth harmonic have an inverse effect on the Kansei value for "clearness-unclearness" in this model. We perceive a musical timber as "beautiful" when it has few harmonic components and their sound pressure levels are low. On the other hand, a musical timber that has strong components higher than the fifth harmonic usually is evaluated as unclear [17][18].

By observing unit *d*, we confirm that the sound pressure levels of the fifth, the sixth and the seventh harmonics determine the Kansei value for "sharpness-

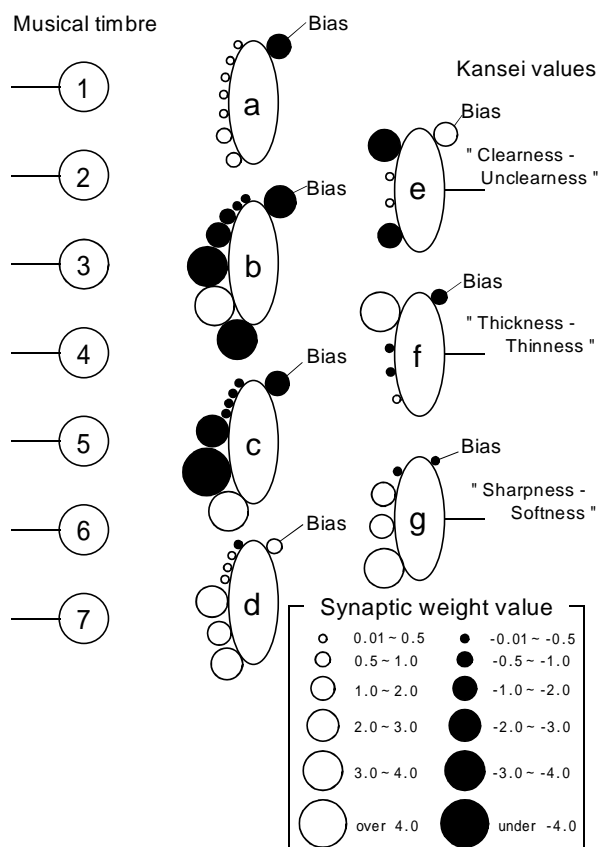


Fig.8 Synaptic weights after training of the network

softness." A musical timber whose high frequency components are emphasized gives us a sharp impression [18]. Furthermore, the values of the sixth and the seventh harmonics which are extracted by units *b* and *c* respectively with the lateral inhibition property influence the Kansei value for "sharpness-softness." The human auditory nervous system has a particular neuron group in the inferior colliculus which analyzes the difference between the frequencies of a sound with lateral inhibition [15][16][17].

The network model in this study has acquired the ability to simulate aspects of human beings which are shown as a result of previous psychology and physiology research.

7. CONCLUSION

In this study, we constructed a model which outputs Kansei values from the input of musical values by using a neural network. The output from this model approximates the response of human Kansei towards musical timber. Furthermore, by analyzing the network model after training, we found that the model shows the human Kansei for musical timber as follows.

- (1) Each harmonic equally influences the Kansei value for "thickness-thinness."
- (2) Harmonic components which are higher than the fourth harmonic have an inverse effect on the Kansei value for "clearness-unclearness."
- (3) Two neurons that have a lateral inhibition property exist in the hidden layer, and these neurons extract the sixth and seventh harmonics. Another neuron in the hidden layer extracts the fifth, sixth and seventh harmonics. These three neurons determine the Kansei value for "sharpness-softness."

This study is based on musical timbre in its narrow sense. The musical timbre for input are chosen so that their harmonics are all in phase. Under these conditions, we observed that the harmonic composition of each musical timbre determines the Kansei values.

Other recent research shows that musical timbre is also affected by the change of the sound with time [10][19]. For example, if a characteristic time portion of the sound such as the attack part is removed from the sound of a musical instrument, it becomes difficult to distinguish it from other instruments. In musical timbre, the non-integer harmonic (the noise component) and small fluctuations are important in distinguishing the musical instrument [20].

We are currently developing a Kansei model for musical timbre in the broad sense with a more complicated construction.

8. REFERENCES

- [1] S.Inokuchi, S.Tanabe, "Kansei Information Processing", Ohmsya, (1994), pp.1-11.
- [2] S.Hitotsumatsu, Y.Muraoka, "Kansei and Information Processing", Kyouritsu Publication, (1993), pp.3-26.
- [3] T.Iida, T.Yanashima, "Perception and Science of Human Perception", Ohmsya, (1995), pp.1-26.
- [4] Y.Matsubara, M.Nagamachi, "Hybrid Kansei Engineering System and Design Support", International Journal of Industrial Ergonomics vol.19, No.2, (1997), pp.81-92.
- [5] S.Ishihara, K.Ishihara, M.Nagamachi, Y.Matsubara, "An Analysis of Kansei Structure on Shoes Using Self-organizing Neural Networks", International Journal of Industrial Ergonomics vol.19, No.2, (1997), pp.93-104.
- [6] G.Onishi, I.Kimura, H.Yamada, "A Sensibility Model for Musical Chords Constructing the Structure of Auditory Sense", Transaction of the Japanese Society of Mechanical Engineers, Vol.65, No.635, (1999), pp. 2787-2793.
- [7] "IEEE Standard Dictionary Electrical and Electronics Terms", IEEE Standards.
- [8] S.Nanba, "Definition of Timbre", Journal of the Japanese Acoustical Society, Vol.49, No.11, (1993), pp.823-831.
- [9] J.Arai, H.Ikeda, "Dictionary of Musical Terms", Rittor Music, (1993), p.208.
- [10] Y.Okura, "The Basic Knowledge of Sound and Music", Kokusyo Kankou Kai, (1999), pp.70-72, 81-82, 90-93.
- [11] R.Plomp, "The Ear as a Frequency Analyzer", Journal of the Acoustical Society of America", Vol.36, (1964), pp.1628-1636.
- [12] N.Anzai, "Know-how of a Synthesizer", Rittor Music, (1996), p.13.
- [13] M.Zukawa, K.Endou, N.Shigeki, "All about the Sound Quality", Seibundou-Shinkou, (1981), pp.65-68.
- [14] N.Nanba, S.Kuwano, "Psychological Measurement for Sound Evaluation", Korona, (1997), pp.107-133.
- [15] T.Ifukube, "Sound-based Assistive Technology for the Disabled", Korona, (1998), pp.9-14.
- [16] T.Ishikawa, "Neurophysiology", Ishiyaku Pub, (1998), pp88-89.
- [17] T.Fukuda, "The Biological Information System", Industry book, (1995), pp.94-109, 122.
- [18] O.Kitamura, M.Sasaki, Building Kansei in Hearing, (1996), pp.35-39, Ongakunotomo.
- [19] Diana Deutsch, "The Psychology of Music", Academic Press, (1982).
- [20] Y.Andou, "Acoustics of Musical Instruments", Ongakunotomo, (1996), pp27-30.