An Operation Plane Using a Neural Network for Intuitive Generation of Robot Motion

Junki Ito*, Masayoshi Kanoh†, Reona Arisawa†, Tsuyoshi Nakamura‡ and Takanori Komatsu§
*Graduate School of Computer and Cognitive Sciences, Chukyo University,
101 Tokodachi, Kaizu-cho, Toyota 470-0393, Japan.
†School of Information Science and Technology, Chukyo University,
101 Tokodachi, Kaizu-cho, Toyota 470-0393, Japan.
‡Graduate School of Engineering, Nagoya Institute of Technology,
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan.
§ Faculty of Textile Science and Technology, Shinshu University,
3-15-1, Tokida, Ueda 386-8567, Japan.

***Abstract*—An operation plane is developed that greatly simplifies the task of editing motion of humanoid robots. The plane uses onomatopoeias, which are words that mimic the appearance or sound of things to produce richly realistic expressions. To create the plane, first, features of known motions for which there are onomatopoeias are extracted by P-type Fourier descriptors. Second, the similarity relationship between the features is learned by using a five-layer auto-associative neural network. Finally, the network's third layer, which has two units, is used as an operation plane. Using this plane, even people who are unfamiliar with robotics can edit motion of humanoid robots intuitively.**

I. Introduction

In recent years, humanoid robots have become a familiar concept due to the extensive amount of research and development work that has been done on them. Humanoid robots can move like human beings, because their physical construction is similar to that of human beings. However, since humanoid robots generally have a lot of degrees of freedom (DOFs), creating/editing their motion is difficult and takes a considerable amount of time.

This paper describes an operation plane using onomatopoeias, which are words that mimic the appearance or sound of things to produce richly realistic expressions. Since onomatopoeias are used to describe the sounds and states of objects, they make it possible to express delicate and realistic representations and subtle differences in a way that cannot be achieved by using ordinary words. This suggests that using onomatopoeias in the editing of robot motion may help to understand sensuously and reduce cognitive burden. With this idea in mind, we have developed an operation plane system in which onomatopoeias are used for editing robot motion (Fig. 1).

In the system, the basic waveform of the robot's movement is first altered by an attribute vector based on the phonemic feature of onomatopoeia [1]. The waveform characteristic variation is then extracted by a P-type Fourier descriptor [2], [3]. Finally, an operation plane that learned the similarity relationships between onomatopoeias is constructed by an auto-associative neural network using the extracted waveform characteristics. When any point on the operation plane is selected, the waveform characteristic of the P-type Fourier descriptor is reproduced with the neural network. Using the characteristic enables the robot to be operated.

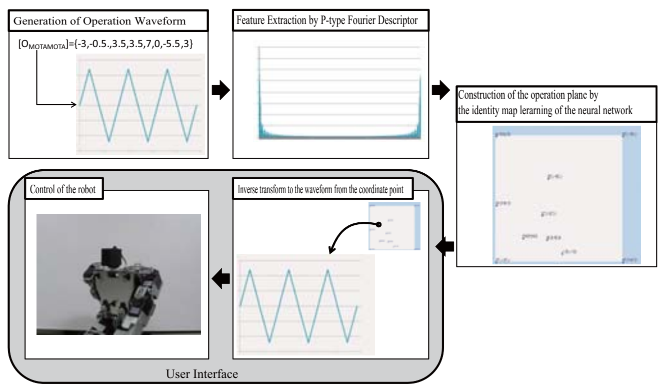


Fig 1. System overview

II. Numerical Modeling of Onomatopoeia

In this paper, we limit the discussion to the $XYXY$ type onomatopoeia.

First, each vowel and consonant in the onomatopoeia is assigned an 8-dimensional value that comprises the values of the attributes, "Strength", "Hardness", "Humidity", "Smoothness", "Roundness", "Elasticity", "Speed", and "Warmth". Note that the value of each attribute is set in the range of $[-2, 2]$.

Take consonant K for example. Its attribute value is defined as follows:

$$K = \{2, 2, 1, 0, 0, 0, 2, -1\}. \tag{1}$$

In the same way, each vowel and consonant in the onomatopoeia is assigned an attribute value, and the onomatopoeia's attribute value is determined by combining these values. Consider for example "$KASAKASA$". It contains four consonants and four vowels, i.e., two "K" consonants, two "S" consonants, and four "A" vowels. The phonemes in

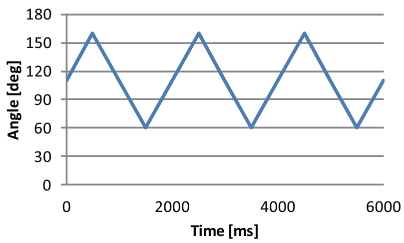
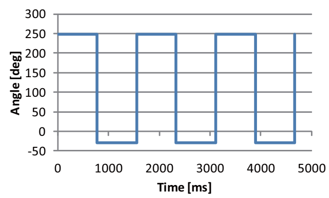


Fig 2. Basic waveform



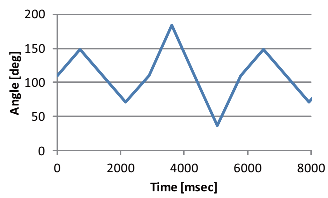


Fig 3. Waveform of "$GASIGASI$" Fig 4. Waveform of "$UROURO$"

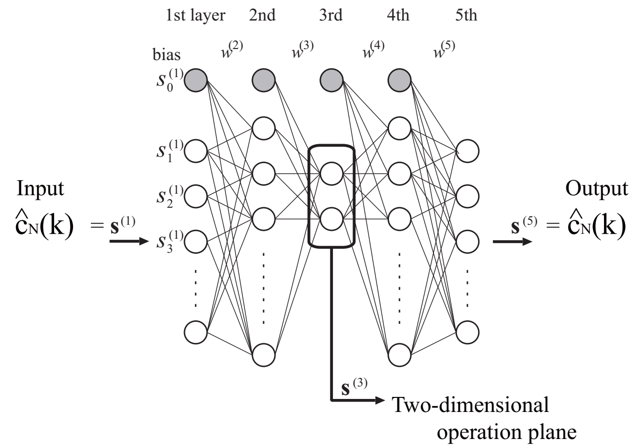


Fig 5. Auto-associative neural network

the $XYXY$ onomatopoeia are described as $X^{(v)}$,$Y^{(v)}$, $X^{(c)}$, and $Y^{(c)}$, where $X^{(v)}$ and $Y^{(v)}$ are the vowels and $X^{(c)}$, and $Y^{(c)}$ are the consonants. When expressed in this way, O_i, the i-th attribute value of onomatopoeia O, is calculated by means of the following equation [1]:

$$O_i = 2X_i^{(c)} + X_i^{(v)} + \frac{(2Y_i^{(c)} + Y_i^{(v)})}{2}. \tag{2}$$

In this equation, $Z_i^{(s)}$, where $Z \in \{X, Y\}$ and $s \in \{v = vowel, c = consonant\}$, represents the i-th attribute value.

III. Waveform Conversion

Using the attribute value computed by means of equation (2), we alter the amplitude and cycle of the basic waveform (Fig. 2) and generate a waveform that represents the motion of onomatopoeia O.

As examples, Figs. 3 and 4 show the waveforms of "$GASIGASI$" and "$UROURO$". For further details on the waveform conversion method, see reference [1].

IV. Waveform Feature Extraction

A P-type Fourier descriptor is used to extract the characteristics of the onomatopoeia's waveform. This descriptor is a kind of discrete Fourier transform. It has a complex function that has a total curvature function in the exponent of the exponential function, as the target of the expansion function.

The descriptor's characteristics are as follows: (1) the endpoints of the original curve and the reproduction curve always match; (2) the endpoints remain unchanged even after translation, expansion, or shrinking of the curve; (3) the more the order used to play the curve is increased, the more the reproduction curve visually approaches the original curve; (4) since the pattern class information of the original curve contains very few low-frequency components, it serves well as a feature quantity parameters. Taking (3) and (4) into consideration, we decided to use a P-type Fourier descriptor as the feature of the robot motion waveform.

Consider $z(j)$, the point sequence of $n+1$ lined at regular intervals of δ on the complex plane. $z(j)$ is expressed by $z(j) = x + iy$ (i is the imaginary unit). $\omega(j)$is determined by the following equation:

$$\omega(j) = \frac{z(j+1) - z(j)}{\delta} \quad (j = 0, ..., n-1), \tag{3}$$

where $\omega(j)$ means P representation of the line shape created by connecting adjacent points $z(j)$ and $z(j+1)$.

$$\hat{c}(k) = \begin{cases} c(k) & (k = 0, 1, ..., n/2) \\ c(n+k) & (k = -n/2+1, ..., -1) \end{cases} . \tag{4}$$

A P-type Fourier descriptor enables the form outline of the shape to be expressed by low-frequency components. In this paper, $\hat{c}_N(k)$ using only low-frequency components for which $|k| \leq N$ is used as the feature value of motion waveform.

$$\hat{c}_N(k) = \begin{cases} \hat{c}(k) & (-N \leq k \leq N) \\ 0 & (otherwise) \end{cases} . \tag{5}$$

Using this as the input signal and teacher signal of the neural network, we construct an auto-associative neural network.

V. Construction of Operation Plane using Neural Network

Using the waveform characteristic quantity obtained by the method described in the previous section as teacher signal input to the neural network, we constructed the operation plane using an auto-associative neural network.

Auto-associative neural network is composed of five layers, and its structure is such that the number of units in the third layer is less than that for input-output (Fig. 5). In this paper the network has two units in the third layer, and their output is used to construct a two-dimensional feature space. This space is used as the operation plane.

Fig 6. KHR-2HV

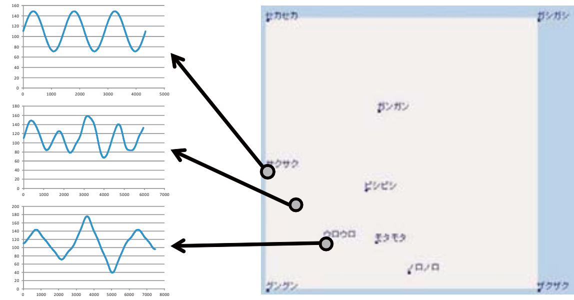


Fig 7. Constructed operation plane and output example

VI. Motion Waveform Extraction from Operation Plane

Motion waveform is extracted by inputting coordinate (x, y) obtained by clicking on the operation plane up to the third layer of the neural network and taking out feature quantity $c_N(k)$ from the fifth layer. $\omega_N(j)$ made using only the low-frequency components for which $|k| \le N$ is called the P representation of the order N of the shape line, and is represented by the following equation:

$$\omega_N(j) = \sum_{k=-N}^{N} \hat{c}_N(k) \exp(-2\pi i \frac{jk}{n}). \tag{6}$$

The curve produced by inverse Fourier transform of $\omega_N(j)$ is called the reproduction curve of the order N.

$$z_N(j) = z_N(0) + \delta \sum_{r=1}^{j-1} \omega_N(r). \tag{7}$$

In this paper, robot motion is created by using this curve.

VII. Evaluation Experiment

A. Robot and Motion Used in Experiment

This evaluation experiment was performed using the KHR-2HV (Fig.6) manufactured by Kondo Kagaku Co., Ltd. This robot has 17 DOFs. The experiment was performed to evaluate whether they could use our method to operate the walking motion intuitively. In this experiment, walking motion is simulated by having the subjects swing the robot's arms.

B. Construction of Operation Plane for Evaluation

We constructed an operation plane for working the robot arms in an anteroposterior direction. Onomatopoeias used to build the operation plane are selected by using the following procedure. Ten subjects took part in the experiment.

First, 20 waveforms that have been granted onomatopoeias was generated, and the robot actually was made to work using these waveforms. They were then asked to answer a five-point Likert-scale type questionnaire to determine whether they felt the onomatopoeias were representative of the robot's movements. From the survey results we found that 10 onomatopoeias earned an average of 3.5 points or more. We then used the waveform characteristic $\hat{c}_N(k)$ to construct the operation plane. The plane is shown in Fig.7.

Figure 7 shows as an example the waveforms generated when the onomatopoeias "$SAKUSAKU$" and "$UROURO$", as well as the waveform reproduced between these two are clicked on. It can be seen that the latter combines the features of the former.

C. Evaluation method

The subjects used the constructed plane to operate the robot and evaluate the motions that were reproducted.

The evaluated onomatopoeias were those used to construct the operation plane and to regenerate the waveforms between those of "$GASIGASI$" and "$GANGAN$", "$ZAKUZAKU$" and "$NORONORO$", and "$BISIBISI$" and "$UROURO$". As before, a five-point Likert-scale type questionnaire was used for the evaluation. We also evaluated the system's usability by using SUS (System Usability Scale) [4] which is a well-known usability evaluation method. With it, usability is investigated simply by having the subject choose a point score from one to five for 10 SUS items. Table I shows the items used in SUS. The total SUS score is calculated by the following equation:

$$SUS = 2.5 \times \sum_{i=1}^{5} \{(x_{2i-1} - 1) + (5 - x_{2i})\}, \tag{8}$$

where x_i is the score for question i. The scores range between 0 and 100, and the higher the score is, the higher is the usability the system is judged to have. Twenty subjects took part in the experiment.

VIII. Experimental Results and Discussion

Figure 8 show the results obtained for 10 onomatopoeias used in the learning, and Fig. 9 shows those for the waveforms generated between a pair of onomatopoeias.

First, let us consider the mapping results obtained for the operation plane of onomatopoeias used in the learning. Some onomatopoeias were evaluated lower in the second questionnaire than they were in the first, but not to any significant degree. This indicates that the motion waveform characteristic

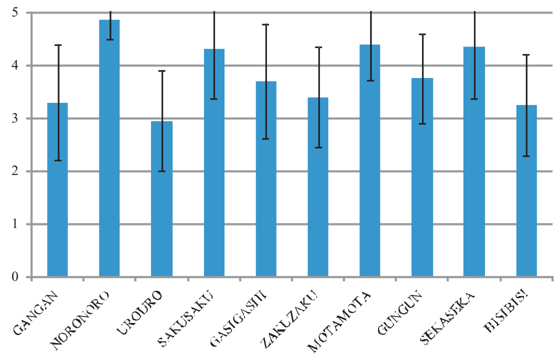


Fig 8. Evaluation results for onomatopoeia used for learning

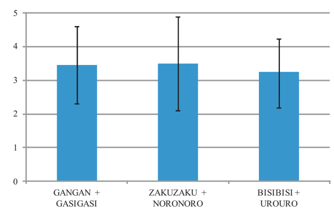


Fig 9. Evaluation results for waveforms generated in intermediate regions

Table I
SUS ITEMS

1	I would like to use this system often.
2	This system is unnecessarily complex.
3	This system is easy for me to use.
4	I may need help from technical experts to use this system.
5	This system's functions are well organized.
6	There is no consistency in this system.
7	I think most users could learn how to use this system easily.
8	This system is very difficult to use.
9	I can use this system with confidence.
10	I would need to learn a lot of things before I could use this system.

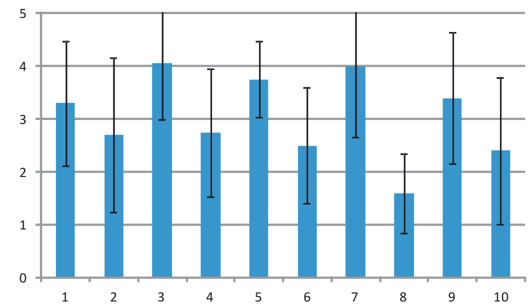


Fig 10. Scores for each item of SUS

was properly extracted by using the P-type Fourier descriptor and identity mapping learning.

Next, let us consider the motion generated between the onomatopoeia pairs on the operation plane. The scores obtained for these were 3.0 points or more, almost the same as those obtained for the onomatopoeias used in the learning. Compared to the latter, however, for the former there was greater variation in the subjects' evaluations. This suggests that the evaluation of waveforms obtained from the operation plane depends on the subject. We also conducted a survey to determine whether there was a difference between the scores for the waveform generated in the middle area of "$BISIBISI$" and "$UROURO$" on the operation plane, which is a dense part, and those for the waveforms generated in the middle areas of "$GANGAN$" and "$GASIGASI$" and "$ZAKUZAKU$" and "$NORONORO$", which are sparse parts. The results showed there were no significant differences in the scores. This suggests that sparse parts on the operation plane are not relevant to the generation of operations, i.e., the motion adapted to relatively the user's impression can be generated in any area on the plane.

Finally, let us consider the SUS results. Figure 10 shows the score obtained for each SUS item. The average score for all items is $66.4(\pm 14.9)$, which indicates this system performs well enough to be put to practical use. The subjects were given an oral interview after the questionnaire, from which we gathered their opinions on the types of scenarios to which the system could be applied. This may have caused them to lower their SUS scores somewhat.

On the other hand, the results showed that the subjects felt it was easy to learn how to use the system and that it was easy to use. We assume that even higher ratings may be obtained if the purpose of the system becomes clearer, i.e., if editing the behavior of robots becomes a more familiar task for more people.

IX. Conclusion

We have developed an operation plane system in which onomatopoeias are used for editing robot motion. The operation plane is constructed by means of a neural network, and using waveforms output from the plane enables even people not familiar with robot operations to easily edit robot motions. We consider that this system has excellent potential to reduce the cognitive burden of motion editing.

In constructing the system, we focused on arm movements in the anteroposterior direction as a part of walking behavior. However, it is possible that this does not allow users to fully express their impressions of onomatopoeias because the range of anteroposterior arm movements is limited and the number of possible movements is small. A subject for future work will be to improve the system by including the movements of joints such as the elbows and ankles to allow for more intuitive system operation.

References

[1] T. Komatsu and H. Akiyama: Expression System of Onomatopoeias for Assisting Users' Intuitive Expressions, Trans. of IEICE A, Vol.J92-A, No.11, pp.752–763, 2009. (in Japanese)

[2] Y. Uesaka: A New Fourier Descriptor Applicable to Open Curves, Trans. of IEICE A, Vol.J67-A, No.3, pp.166–173,1984. (in Japanese)

[3] M. Kanoh, S. Kato and H. Itoh: Efficient Joint Detection Considering Complexity of Contours, Lecture Notes in Artificial Intelligence, Vol.1886, pp.588–598, 2000.

[4] J. Brooke: SUS: A "Quick and Dirty" Usability Scale, P. W. Jordan et al. (eds.), Usability Evaluation in Industry, pp. 189–194, Taylor & Francis, London, UK, 1996.