Frequency-based Skill Analysis for Motion Pictures

Toshiyuki MAEDA[1], Masumi YAJIMA[2], and Akiyoshi WAKATANI[3]
[1] Faculty of Management Information, Hannan University, Japan
[2] Faculty of Economics, Meikai University, Japan
[2] Faculty of Intelligence and Informatics, Konan University, Japan

Abstract— *This paper addresses sports skill analysis using motion pictures, focused on volleyball. In this paper, volleyball play is analyzed with motion picture data recorded by hi-speed cam-coder, where we do not use physical information such as body skeleton model, and so on. Time series data are obtained from the motion picture data with marking points, and analyzed in terms of motion frequency using Mel-frequency cepstral coefficients (MFCCs) as well as Fast Fourier Transform (FFT). We examin to classify individual skills of volleyball attacks.*

Keywords: Motion Picture; Skill Analysis; Motion Frequency; FFT; MFCC

1. Introduction

Many researches for skills, not only engineering skills but also sports skills, treat body structure models and/or skeleton structure models obtained from physical information such as activity or biomechatronical data [1].

Those might be because those researchers believe that internal models of technical skill are structured physically, with some skill levels which are human intention, environmental adjustment, and so on [2].

For instance, Matsumoto and others describes skills of craftsmen, which have actually internal models of structured skill architecture and they choose an action process from internal models adjusted with environment [3]. It is even though hard for skilled workers to represent internal models by themselves. They reflect involuntarily their own represented actions, and achieve highly technical skills with internal models.

We had, however, researched that fore-hand strokes of table tennis play exemplify sports action, and classify skill models using motion picture data analysis without body structure model nor skeleton structure model. We had evaluated those into three play levels as expert/intermediate/novice, and classify the models using data mining technologies [4], [5]. We furthermore had a trial to apply our research framework to other sports skills, such a personal sports skill identification using time series motion picture data, focused on volleyball [6].

There are, on the other hand, some works about human postures. For instance, Nomura et al. [7] present that ground reaction force during human quiet stance is modulated synchronously with the cardiac cycle through hemodynamics. This almost periodic hemodynamic force induces a small disturbance torque to the ankle joint, which is considered as a source of endogenous perturbation that induces postural sway. They consider postural sway dynamics of an inverted pendulum model with an intermittent control strategy, in comparison with the traditional continuous-time feedback controller. They examine whether each control model can exhibit human-like postural sway, characterized by its power law behavior at the low frequency band, when it is weakly perturbed by periodic and/or random forcing mimicking the hemodynamic perturbation. They indicate that the continuous control model with typical feedback gain parameters hardly exhibits the human-like sway pattern, in contrast with the intermittent control model. Further analyses implicate that deterministic, including chaotic, slow oscillations that characterize the intermittent control strategy, together with the small hemodynamic perturbation, could be a possible mechanism for generating the postural sway.

This may implicate that a similar postural control exists even in sports skills, such as expert skills have low frequency motion rather than novice skills. In this paper, we analyze sports skills as for motion frequency using time series motion pictures.

2. Related Works

Wilkinson [8] describes that qualitative skill analysis is an essential analytic tool for physical educators and refers to a process in which a teacher identifies discrepancies between the actual response observed and the desired response. Providing instruction for preserving teachers regarding how to recognize errors has been largely neglected in teacher preparation. The purpose of this study was to evaluate an alternative approach for teaching qualitative skill analysis to undergraduates. The study evaluated the effectiveness of a visual-discrimination training program. The subjects were 18 undergraduate students. The visual-discrimination training program was introduced using a multiple-baseline design across three volleyball skills: the forearm pass, the overhead pass, and the overhead serve. After the introduction of each instructional component, subjects made abrupt improvements in correctly analyzing the volleyball skill. This approach for teaching qualitative skill analysis is one alternative to the conventional techniques currently being used in professional preparation.

Watanabe et al. [9] imply a method for the measurement of sports form. The data obtained can be used for quantitative sports-skill evaluation. They focus on the golf-driver-swing form, which is difficult to measure and also difficult to improve. The measurement method presented was derived by kinematic human-body model analysis. The system was developed using three-dimensional (3-D) rate gyro sensors set of positions on the body that express the 3-D rotations and translations during the golf swing. The system accurately measures the golf-driver-swing form of golfers. Data obtained by this system can be related quantitatively to skill criteria as expressed in respected golf lesson textbooks. Quantitative data for criteria geared toward a novice golfer and a mid-level player are equally useful.

Barzouka et al. [10] examine the effect of feedback with simultaneous skilled model observation and self-modeling on volleyball skill acquisition. 53 pupils 12 to 15 years old formed two experimental groups and one control group who followed an intervention program with 12 practice sessions for acquisition and retention of how to receive a ball. Groups received different types of feedback before and in the middle of each practice session. Reception performance outcome (score) and technique in every group were assessed before and at the end of the intervention program and during the retention phase. A 3 (Group) $\times$ 3 (Measurement Period) multivariate analysis of variance with repeated measures was applied to investigate differences. Results indicated equivalent improvement in all three groups at the end of the intervention program. In conclusion, types of augmented feedback from the physical education teacher are effective in acquisition and retention of the skill for reception in volleyball.

3. Experiment and Discussion

Our research is to identify internal models from observed motion picture data and skill evaluation with represented actions, without measurement of the body structure or the skeleton structure.

3.1 Experiment condition

We here make a target on volleyball attacks as well as the previous research [11], and try further analysis of volleyball attack. For this analysis, we use the same subjects who are six university students.

In terms of skill analysis as representation of plays, we set up two levels as below;

- Expert class: members of university volleyball club, and
- Novice class: inexperienced students.

The subjects are three expert and three novice-level players in this experiment. All subject players are marked at four points as:

- Right elbow
- Right shoulder
- Right waist
- Left knee

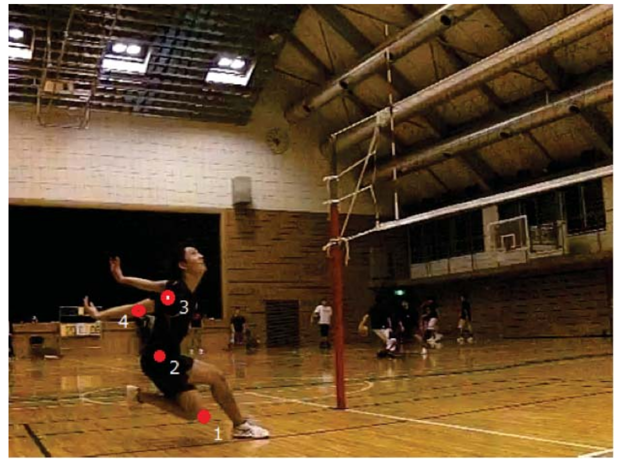

Fig. 1: Measurement markings.

Figure 1 presents the positions of markings for measuring. We record several motion pictures, or movies, for swing traces of attacks by a cam-corder. The recording situation is as follows:

- The cam-corder has a resolution of 512 $\times$ 384 pixels, and a frame-rate of 300 frames per seconds.
- It is installed besides of the players.
- Subjects play in several minutes, and during that time, some attack motions are recorded for each subject player.

3.2 Frequency Analysis using FFT

From the recorded motion pictures, 300 frames are retrieved from the beginning of take-back to the ball until the end of the attack. We have then distributed two dimensional pixel values, as axes positions, of 4 marking points for each frame, where the starting point is set at the shoulder position of the first frame. We firstly analyze frequency of time series motion picture data of each axis using Fast Fourier transform (FFT).

Figure 2 shows frequency spectra of all six subjects for left-knee X-axis. In this figure, n1, n2, and n3 are from the data of novice subjects as well as e4, e5, and e6 are of expert subjects. This figure may roughly implicate novice data tend to have high frequency.

For further analysis, we focus on one novice (n2) and one expert (e4) subjects and present whole spectra of each axis as below, Figure 3,4,5,6,7,8,9,10 respectively.

In those figures, especially figures of 4, 6, 8, 10, namely all y-axes of novice data have more high-frequency data. That implies novice motions may have high frequency motions, though much more investigations are needed for verify this hypothesis.

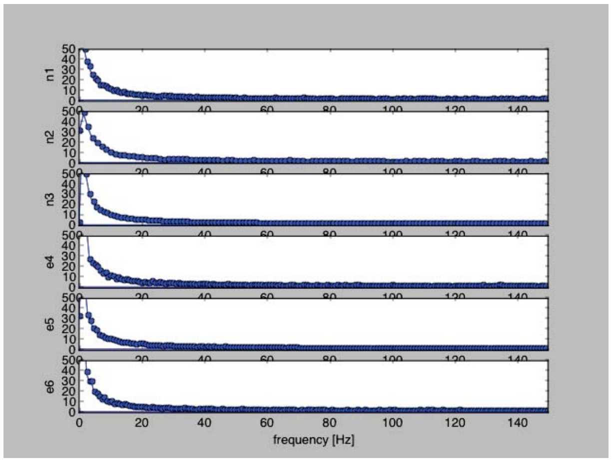

Fig. 2: Frequency analysis for left-knee X-axis of 6 subjects.

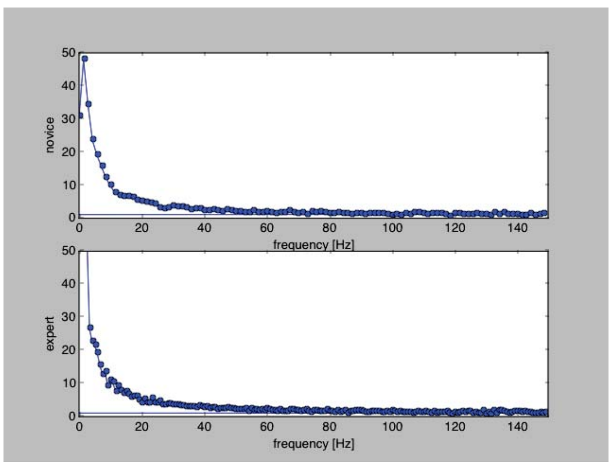

Fig. 3: Frequency analysis for left-knee X-axis.

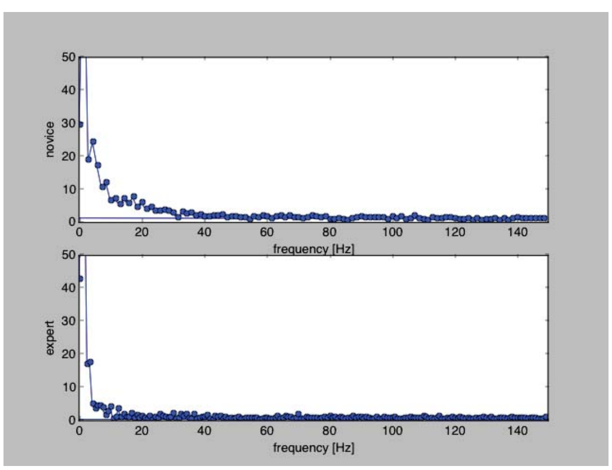

Fig. 4: Frequency analysis for left-knee Y-axis.

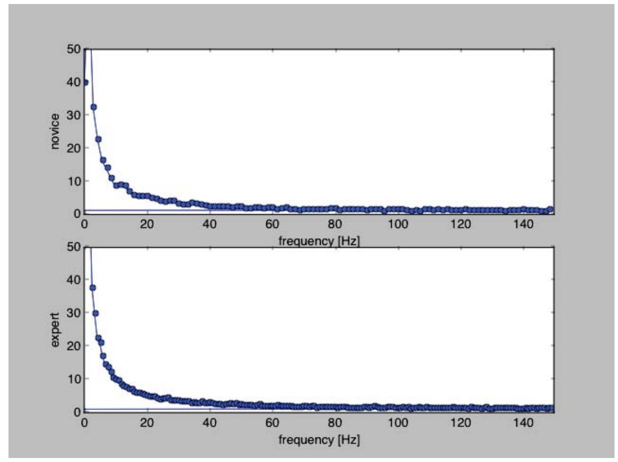

Fig. 5: Frequency analysis for right-waist X-axis.

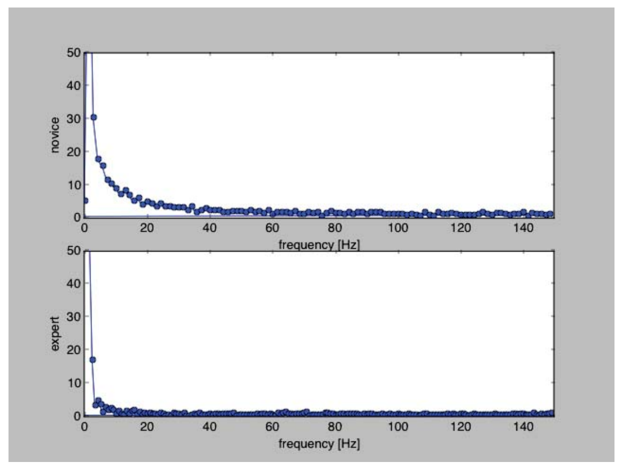

Fig. 6: Frequency analysis for right-waist Y-axis.

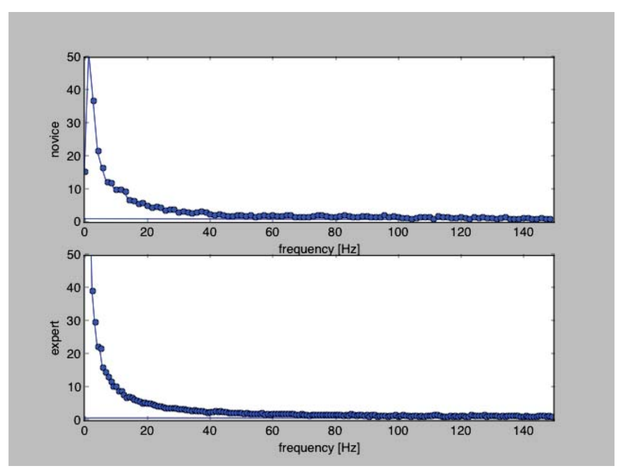

Fig. 7: Frequency analysis for right-shoulder X-axis.

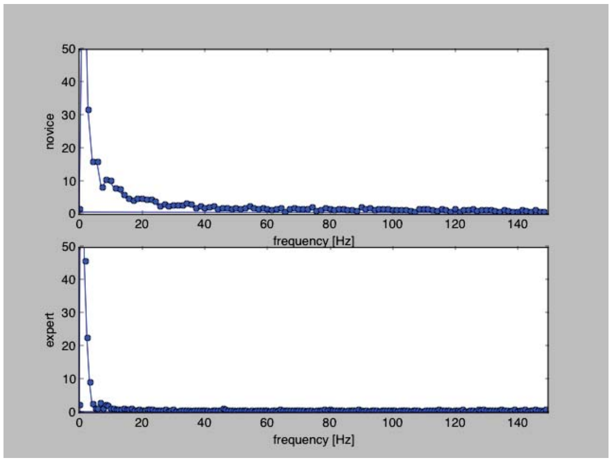

Fig. 8: Frequency analysis for right-shoulder X-axis.

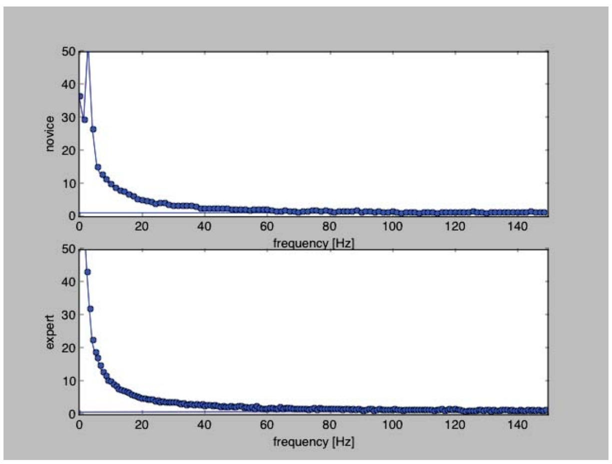

Fig. 9: Frequency analysis for right-elbow X-axis.

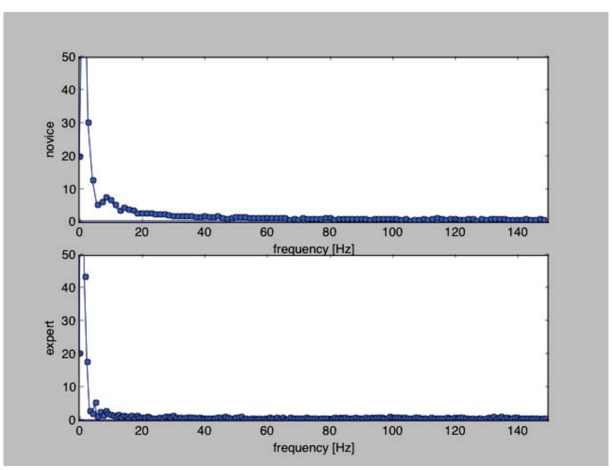

Fig. 10: Frequency analysis for right-elbow Y-axis.

Furthermore, we try to classify frequency data using hierarchical clustering technique, and some have fairly good results. For instance, Figure 11 implies the result of left-knee Y-axis, and this implies novices (n1, n2, n3) and experts (e4, e5, e6) can be clearly classified, though further investigation is needed.

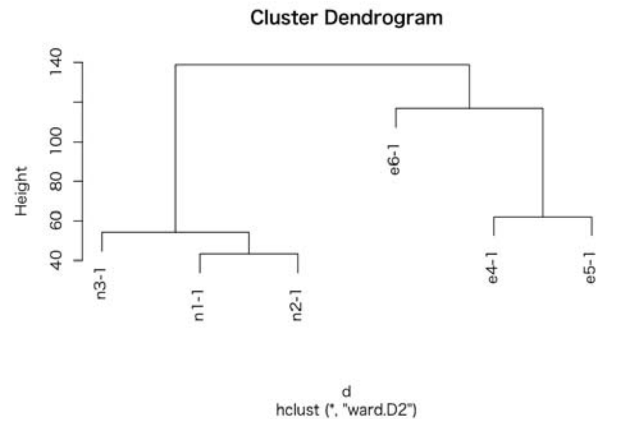

Fig. 11: Clustering analysis for left-knee Y-axis.

3.3 Regression Analysis using FFT and MFCCs

We hence have a trial to classify six subjects to each other. 300 frames (one minute) in the recorded movies are extracted from the start of the take-back till the end of the attack motion as well as previous research. After that, We retrieve two dimensional axes position data of four marking points for each frame indicated as pixel values. The beginning point is fixed, for the basis of the coordinates, at the shoulder position of the first frame. We reconstruct feature vector data from time series data of each axis by Mel-frequency cepstral coefficients (MFCCs) [12], [13] as well as FFT.

The mel-frequency cepstrum (MFC) is representations of the short-term sound power spectrum, and it is based on a log power spectrum with linear cosine transform, and thus on nonlinear mel scale of frequency in the field of sound processing.

MFCCs are coefficients which make up an MFC collectively. Those are derived from type of cepstral representations of audio clips, or a nonlinear "spectrum-of-a-spectrum."

The cepstrum and mel-frequency cepstrum is the most significantly different from the frequency bands in the MFC, and they are equally spaced on the mel scale. That looks after the response of human auditory system more closely than the linearly-spaced frequency bands in the normal cepstrum. For example, this frequency warping may indicate for better sound representation in the field of audio compression.

The Mel scale relates pitch, or perceived frequency, of pure tones to the actual measured frequency [14]. Human being is much better to discern small changes in pitch at low frequencies rather than at high frequencies. To introduce this scale may make the features fit more closely what human being listens.

The formula to convert from frequency to Mel scale is as follows:

$$M(f) = 1125\ln(1+\frac{f}{700})$$

To get from Mel scale back to frequency is as follows:

$$M^{-1}(m) = 700(e^{\frac{m}{1125}} - 1)$$

Brief steps to calculate MFCCs are as follows:

1) Make frames of the signals into short ones.
2) Take the periodogram estimate of power spectra for all frames.
3) Apply the mel filterbank to each power spectrum, then sum the energy for all filters.
4) Calculate logarithms of all filterbank energies.
5) Calculate DCT of all log filterbank energies.
6) Keep DCT coefficients two to thirteen, then discard the rest.

We introduce this methodology in addition to FFT as variation of feature retrieval from motion frequency data.

In the previous research [11], we use hierarchical clustering, but in this research, we try to classify each subject by logistic regression with cross validation. The reason to apply logistic regression to this analysis is because:

- Clustering is a task to group sets of objects by similarity to each other.
- Logistic regression is one of appropriate regression analyses for conduction of dependent variables.
- One of objectives of this research is to investigate the possibility of classification for multi-dependent variables data, and then logistic regression should be fit better than clustering.

At first, we apply the logistic regression classifier to the draft time serial data. Figure 12 implies the classification result as confusion matrix using the draft data.

In this figure, the vertical axis is for "true class", and the horizontal axis for "predicted class." Moreover, 'n1', 'n2', 'n3' are from the data of novice subjects, and 'e4', 'e5', 'e6' are of expert subjects. Densities of squares imply the percentage of correct prediction as indicated in the right side of the figure, and so a denser diagonal square means a better classification result.

This diagram implies the result is quite bad as for classification, though 'e5' may be classified correctly to a certain degree. The accuracy rate is 19.1%.

Figure 13 is the classification result by the FFT-based data.

This diagram implies the result is also bad as for classification, though 'e4' and 'e5' may be classified correctly to a certain degree. The accuracy rate is 23.6%.

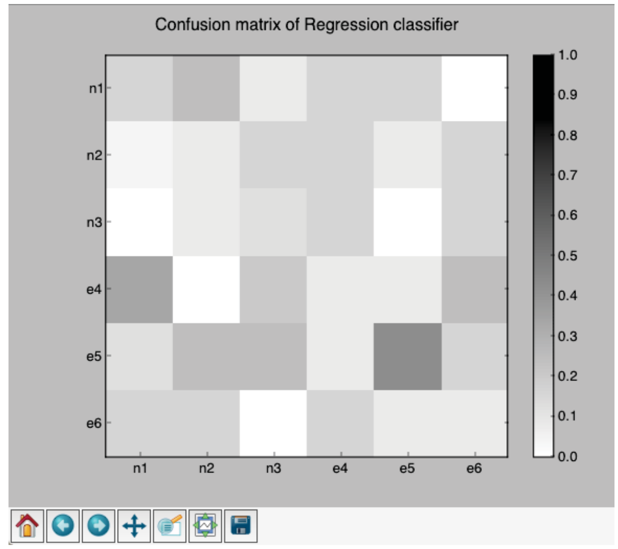

Fig. 12: Confusion matrix of Regression classifier.

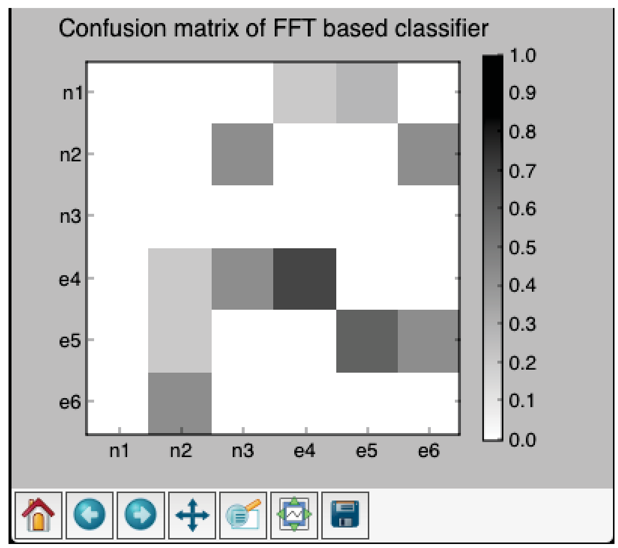

Fig. 13: Confusion matrix of FFT based classifier.

Figure 14 is the classification result by the MFCC-based data. Comparing to the above FFT-based classification, this result does not indicate so good, but a little better than the above result by FFT, as the accuracy rate is 51.6%.

In both cases, 'e5' is classified more clearly rather than the others, and that may imply the skill of 'e5' is quite special, though further experiments have to be carried out as we don't think we have enough amount of data for this sort of analysis.

4. Conclusion

This paper presents frequency analysis as for internal models of skills as evaluation skillfulness for volleyball

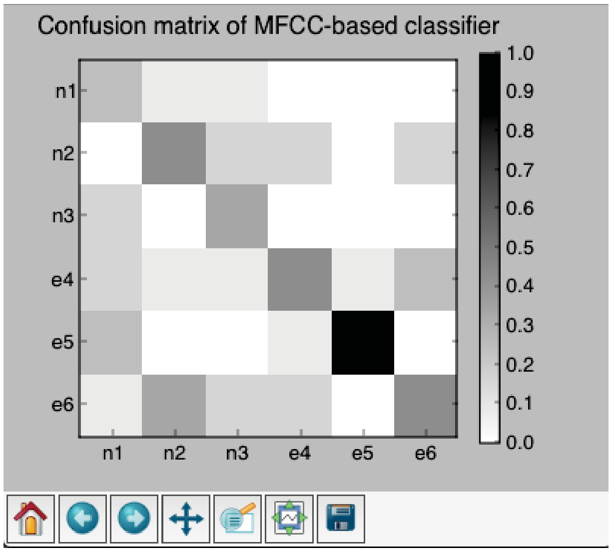

Fig. 14: Confusion matrix of MFCC-based classifier.

attack in terms of motion frequency. We have some experiments and some results show that players can make some categorical groups for technical skills. Moreover, we introduce MFCCs as well as FFT, and we have got some better results that personal analysis may be better categorized, and so on. As future plans, we have to progress further experiments, and measure more precise data and then analyze if needed.

Acknowledgment

Part of this research was supported by JSPS KAKENHI Grant Number 15K02185, 15K03802, and 18K02920. This research was assisted, especially for data management, by Mr. Y. Tamari and Mr. Y. Tsujino at Hannan University. The authors greatly appreciate those.

References

[1] Y. Mochizuki, R. Himeno, and K. Omura, "Artificial skill and a new principle in sports (special issue on digital human : Measurement and modeling of human functions) (in japanese)," *System, Control, and Information*, vol. 46, no. 8, pp. 498–505, 2002.

[2] T. Shiose, T. Sawaragi, K. Kawakami, and O. Katai, "Technological scheme for the preservation of technique from ecological psychological approach (in japanese)," *Ecological Psychology Research*, vol. 1, no. 1, pp. 11–18, 2004.

[3] Y. Matsumoto, *Organization and Skill – Organization Theory of Preservation of Technique (in Japanese)*. Hakuto Shobo, 2003.

[4] T. Maeda, M. Fujii, and I. Hayashi, "Time Series Data Analysis for Sport Skill," in *Proceedings of the 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, Kochi (India), 2012, pp. 392–397.

[5] T. Maeda, M. Fujii, I. Hayashi, and T. Tasaka, "Sport Skill Classification Using Time Series Motion Picture Data," in *Proceedings of the 40th Annual Conference of the IEEE Industrial Electronics Society (IECON 2014)*, Dallas (TX, USA), 2014, pp. 5272–5277.

[6] T. Maeda and M. Yajima, "Personal Skill Identification Using Time Series Motion Picture Data," in *Proceedings of the 41th Annual Conference of the IEEE Industrial Electronics Society (IECON 2015)*, Yokohama (Japan), 2015, pp. 2475–2478.

[7] T. Nomura, S. Oshikawa, Y. Suzuki, K. Kiyono, and P. Morasso, "Modeling human postural sway using an intermittent control and hemodynamic perturbations," *Mathematical Biosciences*, vol. 245, iss.1, pp. 86–95, 2013.

[8] S. Wilkinson, "A training program for improving undergraduates' analytic skill in volleyball," *Journal of Teaching in Physical Education*, vol. 11, iss.2, pp. 177–194, 1992.

[9] K. Watanabe and M. Hokari, "Kinematical analysis and measurement of sports form," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 36, iss.3, pp. 549–557, 2006.

[10] K. Barzouka, N. Bergeles, and D. Hatziharitos, "Effect of simultaneous model observation and self-modeling of volleyball skill acquisition," *Perceptual and Motor Skills*, vol. 104, pp. 32–42, 2007.

[11] T. Maeda and M. Yajima, "Sports skill discrimination with motion frequency analysis," in *Proceedings of The 2016 Future Technologies Conference (FTC 2016)*, San Francisco, CA (USA), 2016, pp. 250–253.

[12] B. L. et al., "Mel frequency cepstral coefficients for music modeling," in *ISMIR*, 2000.

[13] S. Molau, M. Pitz, R. Schluter, and H. Ney, "Computing mel-frequency cepstral coefficients on the power spectrum," in *Proceedings on IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)*, vol. 1. IEEE, 2001, pp. 73–76.

[14] P. Cryptography, "Mel frequency cepstral coefficient (mfcc) tutorial," http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/, 2016.