SELECTED PAPER

Construction of Sound-Quality-Evaluating System Using EEG and GMDH-Type Neural Network

Kiminobu Nishimura¹, Yasue Mitsukura¹, Yohei Tomita² and Akira Hara²

¹ Graduate School of Science and Technology,
Keio University
3-14-1 Hiyoshi, Kohoku, Yokohama City,
Kanagawa 223-8522, Japan
Phone/FAX: +81-45-566-1718
E-mail: nishimura@mitsu.sd.keio.ac.jp
mitsukura@sd.keio.ac.jp

² Foster Electric Company, Limited
1-1-109, Tsutsujigaoka, Akishima City,
Tokyo 196-0012, Japan
Phone/FAX: +81-42-847-3334
E-mail: tomita@foster.co.jp
ahara@foster.co.jp

Abstract

In this paper, we propose a sound-quality-evaluation system using an electroencephalogram (EEG) and a group method of data handling (GMDH) type neural network. EEGs are used in various applications, and we focus on sound quality evaluation. We prepared EEG samples to train a GMDH-type neural network to recognise 3 typical types of sound that were used to create the training data. The results showed that using the GMDH-type neural network improved the recognition rate compared with other methods.

1. Introduction

Recently, loudspeakers have been frequently used in consumer scenes as a result of the development of digital media compression technologies and other multimedia technologies including online digital media stores, which sell music. Additionally, opportunities to listen to music are increasing because of the integration of the music player function in many portable electronic devices, including smart phones and computers. The users select their preferred media and devices, such as loudspeakers, headphones, and amplifiers, to experience music to achieve satisfaction or pleasantness[1]. In other words, users select their devices on the basis of appearance and characteristics. Manufacturers of these devices put a great amount of effort into designing and creating such devices that have the preferred appearances and characteristics determined from a significant number of users. However, it is usually difficult to determine what sound quality of loudspeakers is behind users' preferences. Therefore, various research studies were presented to evaluate them using questionnaires[2]. The results obtained from the questionnaires are usually obscure since the valuation basis of the sound quality is determined by humans. There are differences among individuals, for example, some people care about sound qualities while others do not. Thus, there would be significant errors in evaluation if people do not answer questions

correctly. This error makes the sound-quality-evaluating system useless; hence, an efficient evaluating method is needed. In this research, we define sound quality as a value of pleasantness felt by the listener when the evaluating sound is played. Therefore, we use biological signals, e.g. an electroencephalogram (EEG), to evaluate the sound quality, since EEG is affected by human emotion such as pleasantness. There are several studies on using biological signals, particularly the EEG, which reflects the emotion of the listeners and is being used in musical therapies[3]-[5]. Such research has resulted in systems that measure single- or multiple-channel EEGs and output estimated emotion types, but it is not an effective evaluation because of its discrete property. The ideal output of sound quality can be expressed as numerical values so that the evaluated target can be compared with different loudspeakers. In this paper, we propose a sound-quality-evaluating system with a single-channel EEG to output a numerical value that indicates the sound quality. This paper is organised as follows: in the next section, we explain the existing and proposed methods of analysing EEG. In the third section, we describe our experiment, followed by the results, discussions, and conclusions.

2. EEG Analysis and Proposed Method

Biological signals, such as EEG, can be used to recognise human emotions[6]. There are two types of electroencephalograph for measuring EEG signals. One is the multiple-channel EEG, where a head-gear-type device having multiple channel electrodes positioned referring to the international 10-20 system is used. The other device for measuring EEGs is the single-channel electroencephalograph, which is a hair-band type and can be put on immediately compared with the multiple channel device. Because of its compact body, this device can only measure the Fp1 position of the international 10-20 system. This position is at the left forehead of a human and includes the electric potential caused by neu-

ron activity in the frontal lobe. Since this part of the human brain is related to emotions, the use of only the Fp1 position is sufficient to estimate emotion. The common analysis of EEG is to measure the EEG before and after the task, and calculate the spectrum by applying various frequency analysis methods, for example, fast Fourier transform (FFT), and divide the amplitude of the mean EEG before the task from after the task, for each frequency. The next process would be principal component analysis (PCA), followed by discriminant analysis, for example, Fisher's linear discriminant analysis, to discriminate the different EEGs caused by the different types of sound quality[7]. This conventional method of using PCA is not robust against data that sometimes become irregular. PCA is aimed at acquiring information from the combinations of variables having the most variance, so if some irregular data are used to train the system, it will cause errors leading to overfitting. In this paper, we use a group method of data handling (GMDH) type neural network to perform feature extraction in place of PCA in the conventional method. The GMDH-type neural network uses the "exterior criterion", which will grow a neural network appropriately to prevent overfitting by using an evaluation function. The arguments of this function are the values output when unseen data are set as the input of the network. This will prevent overfitting.

3. Experiment

We prepared three types of sound quality by creating different frequency responses, as indicated in Fig. 1, and applied those equalizers to a specific pair of loudspeakers so the features other than the frequency response will be the same.

Equalizer 1 (EQ1) in Fig. 1 represents a flat response, and EQ3 is a response that cuts below 200 Hz and over 5000 Hz, and boost around 2000 Hz. This sound quality refers to sound from an AM radio. EQ2 represents the sound quality between EQ1 and EQ3. Therefore, in this research, we define good sound quality as the sound quality of the loudspeakers itself, and bad sound quality as a sound quality that of an AM radio. We recruited 32 male and female subjects with ages between 20-50 years old. We instructed the subjects to close their eyes to prevent artifacts in the EEG, and measured EEG for 30 seconds of rest followed by 120 seconds of listening to music with each sound quality, and an additional 30 seconds of rest. We asked each subject to perform this task 6 times for each sound quality in the same day. We used hair-band-type electroencephalograph for measurement at the Fp1 position of the international 10-20 system, and set the reference electrode at the ear tab. The electroencephalograph measures EEG at a sample rate of 128 Hz. We used the conventional method of extracting EEG as follows. First, we perform fast Fourier transform (FFT) to the entire measured EEG signal. This is a result of the FFT parameters of a 1 s window size with no time

Figure 1: Frequency response for three sound qualities

shift and the frequency output by the transform that decreased to be in the range of only 4-22 Hz. Since we measured 180 seconds of EEG, we obtain 19 degree vectors among the total of 180. This process refers to the conventional method[7]. The next process is to extract the EEG difference between before and after the task by the following equations. The transformed data $\mathbf{X}$ can be described as below.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}(1) \\ \mathbf{x}(2) \\ \vdots \\ \mathbf{x}(T) \end{bmatrix} = \begin{bmatrix} x_4(1) & x_5(1) & \cdots & x_{22}(1) \\ x_4(2) & x_5(2) & \cdots & x_{22}(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_4(T) & x_5(T) & \cdots & x_{22}(T) \end{bmatrix} \quad (1)$$

In the equation above, $x_f(\tau)$ indicates the amplitude value of f Hz at τ th second. T is the measured time, 180. The divided EEG after by the average of before the task, $\tilde{\mathbf{X}}$ can be calculated by the following equation.

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \\ \vdots \\ \tilde{\mathbf{x}}_{30} \end{bmatrix} = \begin{bmatrix} \frac{x_4(T-29)}{\bar{x}_4} & \frac{x_5(T-29)}{\bar{x}_5} & \cdots & \frac{x_{22}(T-29)}{\bar{x}_{22}} \\ \frac{x_4(T-28)}{\bar{x}_4} & \frac{x_5(T-28)}{\bar{x}_5} & \cdots & \frac{x_{22}(T-28)}{\bar{x}_{22}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_4(T)}{\bar{x}_4} & \frac{x_5(T)}{\bar{x}_5} & \cdots & \frac{x_{22}(T)}{\bar{x}_{22}} \end{bmatrix} \quad (2)$$

where

$$\bar{x}_j = \frac{1}{30} \sum_{i=1}^{30} x_j(i) \quad (3)$$

Table 1: Recognition rate using PCA

Subject	Recognition Rate [%]	Subject	Recognition Rate [%]
1	44.02	17	44.23
2	41.03	18	39.53
3	42.31	19	47.44
4	40.60	20	44.87
5	40.17	21	39.53
6	40.81	22	41.24
7	41.45	23	49.79
8	49.57	24	44.23
9	39.32	25	42.74
10	42.52	26	41.88
11	34.40	27	50.00
12	41.45	28	44.02
13	45.09	29	44.87
14	43.16	30	48.29
15	44.87	31	38.25
16	47.44	32	40.38
Average		43.23	

Table 2: Recognition rate using GMDH-type neural network

Subject	Recognition Rate [%]	Subject	Recognition Rate [%]
1	64.86	17	55.00
2	68.22	18	61.14
3	73.75	19	62.67
4	61.78	20	66.25
5	61.00	21	60.00
6	61.89	22	62.33
7	57.33	23	63.33
8	53.20	24	63.33
9	54.69	25	66.75
10	63.78	26	62.25
11	61.43	27	74.40
12	64.44	28	63.78
13	66.00	29	55.41
14	55.24	30	68.44
15	68.44	31	58.22
16	56.22	32	64.44
Average		62.05	

Finally, we apply a moving average filter.

$$\tilde{\mathbf{X}}_{\text{SMA}_t} = \frac{1}{5}(\tilde{\mathbf{x}}_t + \tilde{\mathbf{x}}_{t+1} + \tilde{\mathbf{x}}_{t+2} + \tilde{\mathbf{x}}_{t+3} + \tilde{\mathbf{x}}_{t+4}), (t = 1, 2, \dots, 26) \quad (4)$$

We now have 26×19 matrix data. As a consequence, there are 26 usable data for each EEG measurement. We then construct a GMDH-type neural network and train this network using the prepared data. This network is grown automatically by its method creating layers, and the polynomial function of the neuron used is set as $f(\omega, \mathbf{u})$.

$$f(\omega, \mathbf{u}) = \omega_1 + \omega_2 u_1 + \omega_3 u_2 + \omega_4 u_1 u_2 + \omega_5 u_1^2 + \omega_6 u_2^2 \quad (5)$$

where ω_i is the weight of the i th term and u_j is the j th input of the neuron ($j = 1, 2$). We also stipulate that each neuron can take an input from the input layer in the network. The network will only have one output neuron, which will be a numerical value. The target signals were set to -1, 0, and 1 for EQ1, EQ2, and EQ3, respectively. The evaluation function to investigate the neuron is calculated by the least-square methods using the training data and its target signal. The specific method used for this is Levenberg-Marquardt algorithm, which is said to yield good results for nonlinear problems. For comparison, we applied PCA to these data and selected the first two principal components.

4. Results and Discussion

We calculated the recognition rates of the 3 sound qualities by analysing the EEG signals. We used the threshold method and Fischer linear discriminant analysis for feature

extraction by the GMDH-type neural network and PCA, respectively. 10-fold cross validation was adopted for both the PCA and GMDH-type neural network methods, and the results of the final values are indicated as the recognition rates in Table 1 and Table 2, respectively. The bold font items in both tables represent the highest score among all subjects. We can see that the method of using the GMDH-type neural network improved the recognition rate for all subjects. Fig. 2 indicates the plots for the data obtained by the feature extraction described in the previous section and by applying PCA, selecting the first two principal components. Both methods used the same data. Dots, circles, and squares describe data of EQ1, EQ2, and EQ3, respectively. Figure 3 indicates the histogram for the output of the neural network. The output values fall in the range from -1 to 1, because the target signal were set as -1, 0, or 1. The stripe and shadow patterns in the legend in Fig. 3 indicate the data for EQ1, EQ2, and EQ3. The horizontal axis is the output value of the network, and the vertical axis is the count of data within the range of output values. The covered area is shown in the mixed colours indicated in the legend. We set the thresholds for recognising EQ1, EQ2 and EQ3 as -0.5 and 0.5. Results in both Figures 2 and 3 are for only subject 3. This subject showed a significant difference between the two methods: we can see a large overlapped area of multiple colour plots in Fig. 2, but the histogram in Fig. 3 is clearly separated. This resulted in a recognition rate difference in Table 1 and Table 2 that is the highest among all subjects. We could not put all plots and histograms in the limited paper area, but it was common that the principal component plot covered large areas compared

Figure 2: Principal component plots for subject 3

with that of the GMDH-type neural network method that had smaller covered areas, which causes misrecognition.

The results showed that the method of using the GMDH-type neural network is better than the method of using PCA. This can be a result of the appropriate selection of frequency and their weights to separate the output values. PCA is a method for identifying the axis that has the largest variance, but it does not result in enough separation to enable discrimination by a simple linear analysis. However, we have not yet investigated the actual cause of the differences of sound quality, by frequency nor the cause of misprediction when GMDH-type neural network method is used. The proposed method of using the GMDH-type neural network outputs a one-dimensional value, compared with the two-dimensional values when PCA was used. There also is a need to assess what the output value of the neural network actually means, which should be pleasantness as we are defining the sound quality.

5. Conclusions

In this paper, we proposed an EEG analysis method of using a GMDH-type neural network to evaluate the sound quality of a pair of loudspeakers. We applied frequency transformation, the same as in the conventional method, but by creating the GMDH-type neural network that outputs one value to extract the EEG difference between sound quality, we improved the separation of feature scores. The three types of equalizers followed linearity in the output of the proposed system, as the actual sound qualities were also defined linearly. In our future work, we will investigate a system to be actually used to qualify loudspeakers, by testing many different types of loudspeaker.

Figure 3: Histogram using GMDH-type neural network for subject 3

References

- [1] A. Neuman, M. Bakke, C. Mackersie, S. Hellman and H. Levitt: The effect of compression ratio and release time on the categorical rating of sound quality, *J. Acoust. Soc. Amer.*, Vol. 103, No. 5, pp. 2273-2278, 1998.
- [2] A. Gabrielsson and H. Sjogren: Perceived sound quality of sound reproducing systems, *J. Acoust. Soc. Amer.*, Vol. 65, No. 4, pp. 1019- 1033, 1979.
- [3] T. Ogawa, S. Karungaru, Y. Mitsukura, M. Fukumi and N. Akamatsu: The EEG analysis by using a neural network in listening to music (in Japanese), *IEICE Technical Report* 2007, pp. 5-9, 2007.
- [4] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann and J.-H. Chen: EEG-based emotion recognition in music listening, *IEEE Trans. Biomedical Engineering*, Vol. 57, No. 7, pp. 1798-1806, 2010.
- [5] Y.-P. Lin, J.-R. Duann, J.-H. Chen and T.-P. Jung: Electroencephalographic dynamics of musical emotion perception revealed by independent spectral components, *Neuroreport*, Vol. 21, No. 6, pp. 410-415, 2010.
- [6] S. Sanei and J. Chambers: *EEG Signal Processing*, Wiley-Interscience, 2009.
- [7] M. Nakanishi, Y. Mitsukura and A. Hara: EEG analysis for acoustic quality evaluation using principal component analysis, *Proc. 20th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 321-324, 2011.