Evaluation of auditory impressions induced by HVAC sound using predictive model

Hajime YANO[1,2], Tetsuya TAKIGUCHI[1], Yasuo ARIKI[1], Masaru KAMIYA[3], Seiji NAKAGAWA[4,2]
[1]Graduate School of System Informatics, Kobe University
[2]National Institute of Advanced Industrial Science and Technology (AIST)
[3]DENSO Corporation
[4]Center for Frontier Medical Engineering, Chiba University

Abstract: The evaluation of subjective impressions induced by environmental sounds using neurophysiological indices has been proposed in recent years. In this paper, we focus on the evaluation of HVAC (heating, ventilation and air conditioning) sounds, and models that predict subjective coolness/preference induced by time-varying HVAC sound from brain activities were constructed. First, magnetoencephalographic (MEG) measurements were carried out to measure brain activities while hearing HVAC sound with paired comparison task. Second, feature vectors representing time-frequency components of brain activities on the whole head were extracted from MEG data using the time-frequency analysis and nonnegative tensor factorization (NTF). And third, two kinds of predictive model were constructed from the brain feature vectors and comparative judgments to pairs of stimuli using a regression model or an SVM-based method. Evaluation experiments show that the SVM-based method is more effective than the regression model.
Keywords: Auditory impressions, Magnetoencephalography, HVAC sound, Nonnegative tensor factorization

1. Introduction

To improve the quality of environmental sounds, evaluations of the subjective impressions induced by them are essential. Evaluations of subjective impressions using neurophysiological indices have been proposed in recent years(1-4). It is expected that neurophysiological methods are less influenced by cognitive bias and that subconscious impressions can also be evaluated, against conventional psychological methods.

To create a practical neurophysiological index, relationships between subjective auditory impressions and brain activities have been investigated. Soeta *et al.* reported that the temporal stability of 8-13 Hz cortical magnetic activity, i.e. alpha oscillation, had a positive correlation with subjective preferences of speech sounds with reverberation(1), and a negative correlation with annoyance of bandpass noise(2). Although these reports indicated the possibility of evaluation of auditory impressions of environmental sounds using neurophysiological indices, they have not come to practical indices yet.

On the other hand, creating a neurophysiological index for evaluation of impressions is to find a mapping from brain activities and impressions. Thus, such a mapping can be learned by machine learning methods.

In this paper, we focused on an evaluation of automotive HVAC (heating, ventilation and air conditioning) sounds, which affect comfortableness in a car largely because they continuously make loud noise. First, magnetoencephalographic (MEG) measurements were conducted to record brain activities while hearing each HVAC sounds, as well as paired-comparative judgments to pairs of HVAC sounds on coolness/preference. Second, the time-frequency features of MEG data were calculated. And then, low dimensional features were extracted from multi-dimensional time-frequency feature using nonnegative tensor factorization (NTF). Finally, the models that predict subjective coolness/preference induced by HVAC sound from brain activities were constructed using a regression model or the support vector machine (SVM)-based method, and their performance was evaluated with the prediction accuracies of paired-comparative judgment.

2. Recordings of MEG

Auditory stimuli were generated from the recorded noise of an actual vehicle HVAC system using linear predictive coding (LPC)(5), and these were amplitude modulated sinusoidally to give time variation. A model HVAC sound synthesized using 150 linear predictive coefficients.

Seven stimuli were made by changing the modulation frequency: 0 (without modulation), 0.2, 0.4, 0.6, 0.8, 1.6 and 3.2 Hz. The modulation depth was fixed at 0.15. The duration of stimuli was 5.0 s. The sound-pressure level of stimuli was set at a level where a stimulus at 0 Hz can be heard “clearly and nicely.”

Eight participants, who were 21-24 years old and had normal hearing, took part in the MEG measurements. Necessary information regarding the experiment was given to the subjects, and informed consent was obtained prior to the experiment. The experiment was approved by the Institutional Review Board on Ergonomic Research of AIST.

Four stimuli, including the most- and least-cooled/preferred stimuli in subjective evaluation tests carried out before the MEG measurements, were selected for each subject.

The pair of stimuli were presented to participant's ears by the insertion-type earphones, and each subject was asked to report which of stimuli he/she feel cooler/prefer, i.e. paired-comparison task. The measurements were performed in a magnetically shielded room using a 122-channel whole-head neuromagnetometer (Neuromag-122™, Neuromag Ltd.). Magnetic signals were passed through the analog filter with a passband from 0.03 to 100 Hz and sampled at 400 Hz. Comparative judgments on paired stimuli were also recorded at the same time.

3. Feature extraction

First, the continuous wavelet transform (CWT) was performed on the raw data to obtain the frequency response at each time. The complex Morlet wavelet was used as the mother wavelet. The CWT coefficients of the raw data at each channel were calculated at the time from 0 to 5000 ms in intervals of 100 ms, for each frequency from 4 to 100 Hz by 1 Hz. Because of the CWT coefficients with a complex wavelet are complex numbers, their amplitudes were used as the time-frequency feature of MEG data.

Second, the low dimensional feature extracted from the time-frequency feature using NTF[6]. The time-frequency feature that has I frequency bins and J time samples, obtained from L epochs including K channels signals, are represented as a 4-way tensor with nonnegative elements, $\underline{\mathbf{X}} \in \mathbf{R}_{+}^{I \times J \times K \times L}$. NTF can extract feature preserving multi-dimensional structure of tensorial data. NTF of $\underline{\mathbf{X}}$ is given by

$$\underline{\mathbf{X}} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \circ \mathbf{t}_r, \qquad (1)$$

where $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_R] \in \mathbf{R}_{+}^{I \times R}$, $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_R] \in \mathbf{R}_{+}^{J \times R}$, $\mathbf{C} = [\mathbf{c}_1 \ldots \mathbf{c}_R] \in \mathbf{R}_{+}^{K \times R}$ and $\mathbf{T} = [\mathbf{t}_1 \ldots \mathbf{t}_R] \in \mathbf{R}_{+}^{L \times R}$ denotes basis matrices corresponding to frequency, time, channel and epoch, respectively. The symbol $\circ$ denotes outer product of vectors. $\underline{\mathbf{X}}$ is represented by the sum of R rank-1 tensors, $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \circ \mathbf{t}_r$.

It is considered that a 4-way rank-1 tensor $\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \circ \mathbf{t}_r$ can express a brain activity on specific frequencies, latencies and channels in all epoch. Thus, each row vector of $\mathbf{T}$ represents the intensity of brain activities for each epoch. If $\underline{\mathbf{X}}$ is decomposed into fewer components with smaller R, each row vector of $\mathbf{T}$ can be used as the low dimensional brain feature.

4. Evaluation of impressions using predictive model

Regression models provide a simple way to learn a mapping between feature vectors and scale values. Scale value of impressions for each stimulus was given to a response variable of regression. In this paper, support vector regression (SVR) was used as a regression model.

On the other hand, although regression models assumed that scale values of impressions obtained from psychological tests have fixed relationships, it is possible that comparative judgment on a pair of stimuli is not always consistent in all trial when making judgment of paired stimuli is difficult. To learn change of comparative judgment for each trial, we employed SVM-based method[7]. In this method, a predictive model that classifies the difference between paired feature vectors into two classes corresponding to paired-comparative judgment is trained using the SVM framework. Finally, we solved the following maximization problem:

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \quad \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \mathbf{d}_i, \mathbf{d}_j \rangle, \\ &\text{s.t.} \quad 0 \le \alpha_i \le C, \end{aligned} \qquad (2)$$

where $\mathbf{d}_i$ denotes the difference between paired feature vectors, and C determines the trade-off between the margin and errors of the training data.

In evaluation experiments, both predictive models were constructed for each subject and their performance was evaluated. Paired-comparative judgments were predicted from magnitude relations between two predicted scale values, and prediction accuracies were calculated. Scale values of subjective coolness/preference induced by HVAC sound were calculated from paired-comparative judgment recorded in MEG experiments using Thurstone's method of paired-comparison (Case V). To learn the nonlinear mapping, the radial basis function was used as kernel function in both predictive models.

Table 1 Prediction accuracies of paired-comparative judgment on coolness.

	SVR		SVM-based	
	Closed	Open	Closed	Open
Mean accuracy [%]	63.7	52.8	87.9	56.0

Table 2 Prediction accuracies of paired-comparative judgment on preference

	SVR		SVM-based	
	Closed	Open	Closed	Open
Mean accuracy [%]	65.1	51.3	91.1	57.2

5. Results and discussion

Tables 1 and 2 show mean prediction accuracies on coolness and preference, respectively. In these tables "Closed" denotes evaluation using the training data, and "Open" denotes evaluation using the testing data. In the cases of both coolness and preference, the mean accuracies of the proposed method were higher than ones of SVR. The mean prediction accuracy even with the training data was only 65.1% because, in learning step, the regression model was given not comparative relations between paired stimuli, but scale values that had fixed magnitude relations.

These results indicate that the learning of paired-comparative judgment can improve performance of predictive model for neurophysiological evaluation of impressions with paired-comparison task. However, the prediction accuracy with testing data was less than 60%, therefore additional improvement of feature extraction and predictive model is needed for practical use.

Reference

1. Y. Soeta *et al.*, "Magnetoencephalographic responses corresponding to individual subjective preference of sound field," J. Sound Vib., vol. 258, no. 3, pp.419-428, 2002.

2. Y. Soeta *et al.*, "Magnetoencephalographic responses correspond to individual annoyance of bandpass noise," J. Sound Vib., vol. 277, no. 3, pp. 479-489, 2004.

3. S. Nakagawa *et al.*, "Evaluation of auditory impression of music using brain activity," ICIC Exp. Lett., vol. 7, pp. 1545-1550, 2013.

4. Y. Okamoto *et al.*, "Human cortical activity related to subjective preference for fluctuating light, J. Temp. Design Architect. Environ., " vol. 7, pp. 9-17, 2007.

5. T. Hotehama *et al.*, "Auditory impression of the coolness and warmness of automotive HVAC noise," in Proc. INTER-NOISE 2015, 2015.

6. A. Cichocki *et al.*, "Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation," Willey, 2009.

7. R. Herbrich *et al.*, "Learning preference relations for information retrieval," in Proc. AAAI Workshop Text Categorization and Machine Learning, pp. 80-84, 1998.