GOM-Face: GKP, EOG, and EMG-Based Multimodal Interface With Application to Humanoid Robot Control

Yunjun Nam, Bonkon Koo, Andrzej Cichocki, *Member, IEEE*, and Seungjin Choi*, *Member, IEEE*

***Abstract*—We present a novel human–machine interface, called *GOM-Face*, and its application to humanoid robot control. *The GOM-Face* bases its interfacing on three electric potentials measured on the face: 1) glossokinetic potential (GKP), which involves the tongue movement; 2) electrooculogram (EOG), which involves the eye movement; 3) electromyogram, which involves the teeth clenching. Each potential has been individually used for assistive interfacing to provide persons with limb motor disabilities or even complete quadriplegia an alternative communication channel. However, to the best of our knowledge, *GOM-Face* is the first interface that exploits all these potentials together. We resolved the interference between GKP and EOG by extracting discriminative features from two covariance matrices: a tongue-movement-only data matrix and eye-movement-only data matrix. With the feature extraction method, *GOM-Face* can detect four kinds of horizontal tongue or eye movements with an accuracy of 86.7% within 2.77 s. We demonstrated the applicability of the *GOM-Face* to humanoid robot control: users were able to communicate with the robot by selecting from a predefined menu using the eye and tongue movements.**

***Index Terms*—Electromyogram (EMG), electrooculogram (EOG), glossokinetic potentials (GKP), human–machine interface, multimodal interface.**

I. Introduction

VARIOUS assistive technologies have been developed for persons with limb motor disabilities to help them perform daily tasks. Among these techniques, the interfaces exploiting the bioelectrical potentials from the facial region, such as electrooculogram (EOG), electromyogram (EMG), and glossokinetic potential (GKP, the electrical potential response generated by tongue movement), have attracted considerable attention as promising technologies for the quadriplegic patients. The facial muscles involved in the above signals are directly connected to the brain by the cranial nerves, and the distance from the brain is shorter than the limbs [1]. Therefore, the facial muscles generally escape severe damage in spinal cord injuries and are the last to be affected in most neuromuscular degenerative disorders, such as amyotrophic lateral sclerosis [1], [2].

To utilize this advantage, various assistive technologies based on bioelectrical potentials have been developed. The EOG-based gaze detection interfaces, such as wearable EOG goggles [3] and EOG pointer [4], have been developed for various applications, like wheelchair [5] and visual navigation [6]. Generally compared to vision-based gaze detection techniques, the EOG-based technique can operate in a less obtrusive way by attaching inexpensive, small, and light-weight electrodes (e.g., 3M red dots) on the skin near the eyes without obstructing the user's view [3]. Recently, we also proposed a new tongue–machine interface that uses GKP to drive an electric wheelchair [1]. The proposed interface can detect tongue positions solely by using GKP recorded from outside the mouth so it is more comfortable and hygienic than the techniques that require the injection of the sensors [7] or the magnets [8] inside the mouth.

With these advancing technologies, multimodal interface approaches [9], which process two or more combined user input modes—such as gaze, gesture, and even imagery movement—in a coordinated manner, are also developed to achieve better robustness, accuracy, and usability [10], [11]. For example, Tsui *et al.* integrated EOG and EMG signals to control a wheelchair [12]. In this paper, the EMG signals were used to control the advancing direction of the wheelchair and the EOG signals are used to adjust its speed. Usakli *et al.* suggested a novel multimodal interface that combines an EOG-based function selection and a P300-based speller interface [13].

With the aid of recent research on multimodal interfaces, we propose a novel human–machine interface that is based on three electric potentials: GKP, EOG, and EMG. We named this interface *GOM-Face*, derived from **G**KP, E**O**G, and E**M**G detection inter**face**. The *GOM-Face* can detect two kinds of horizontal tongue movements, two kinds of horizontal eye movements, and the teeth-clenching movement, by analyzing the electric potential signals recorded on the face.

To implement the interface, we need to resolve the interference between EOG and GKP. EOG is the electrical potential response generated by the eye movements [14, pp. 107–109]. When a cornea positively charged with respect to the posterior pole of the eye moves during the eye movements, it changes the

Manuscript received January 5, 2013; revised April 13, 2013, June 19, 2013, and August 26, 2013; accepted August 27, 2013. Date of publication September 5, 2013; date of current version January 16, 2014. This work was supported by the National Research Foundation of Korea (NRF-2010-0018828 and NRF-2010-0018829), NRF World Class University Program (R31-10100). A portion of this work was carried out while Y. Nam was visiting the Lab for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, W Japan. *Asterisk indicates corresponding author*.

Y. Nam and B. Koo are with the School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang 790-784, Korea (e-mail: druid@postech.ac.kr; bkkoo@postech.ac.kr).

A. Cichocki is with the RIKEN Brain Science Institute, Wakō 351-0198, Japan, and also with the Systems Research Institute Polish Academy of Science, 01-447 Warszawa, Polska (e-mail: a.cichocki@brain.riken.jp).

*S. Choi is with the Department of Computer Science and Engineering and Division of IT Convergence Engineering, Pohang University of Science and Technology, Pohang 790-784, Korea (e-mail: seungjin@postech.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2280900

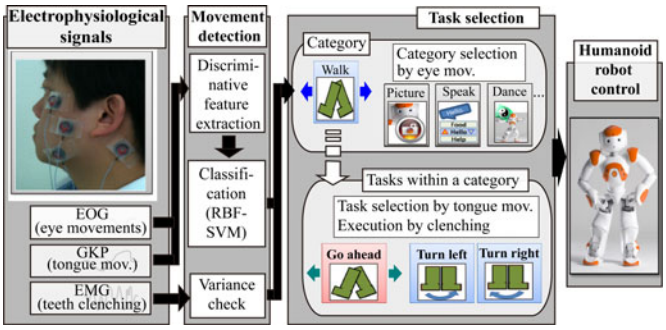

Fig. 1. *GOM-Face* is a novel bioelectrical potential-based human–machine interface that recognizes three kinds of facial movements: 1) eye movements, 2) tongue movements, and 3) teeth-clenching movements. These movements are assigned to a two-level hierarchical interface, so that the subject can select a category and task to order a humanoid robot to perform various actions.

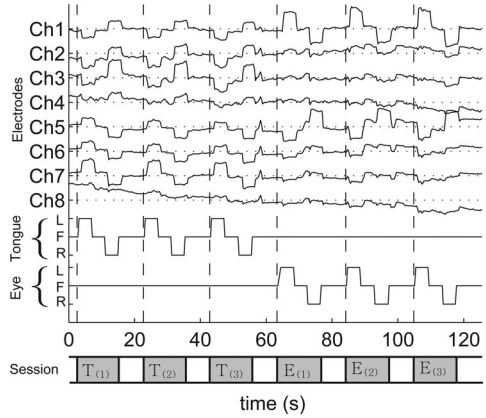

Fig. 2. Recorded potential signals during three sessions of the tongue movements ($T_{(1)}$–$T_{(3)}$) and three sessions of eye movements ($E_{(1)}$–$E_{(3)}$). The given cues for the tongue and eye movements are plotted at the bottom of the figure. The line labeled "Tongue" represents the cue for the tongue movements, and the line labeled "Eye" represents the cue for the eye movements. The tasks (L, F, and R) are denoted by 1, 0, and −1, respectively.

potential levels of the surrounding skin. In the same way, GKP is generated from the tip of the tongue, which has a negative electrical charge with respect to the root [14, pp. 116], [15]. If the tongue moves to the left side while staying in contact with a buccal wall, the potential levels in the left facial region decrease and those in the right region increase, and vice versa. Because GKP and EOG are generated via the same phenomenon (i.e., movement of the charged organ), their spatial patterns of potential response are similar. As a result, GKP from the tongue movements can be mistakenly identified as EOG from eye movements and vice versa.

We resolve this interference problem by searching for discriminative features that are sensitive to the tongue movement but insensitive to the eye movement, and vice versa. To extract the features, we exploit the generalized eigenvalue decomposition of two covariance matrices: a tongue-movement-only data matrix and an eye-movement-only data matrix. This decomposition approach has been successfully applied to electroencephalogram (EEG) analysis [16], particularly for solving motor imagery classifications termed common spatial patterns [17]. It removes the correlation between two kinds of signals by generating a new feature space where two distributions for each signal are maximally dissimilar along the new axes [18]. With these novel features, the developed interface system can detect and discriminate among the four kinds of tongue and eye movements with an average accuracy of $86.7 \pm 8.28\%$ and an average response time of 2.77 ± 0.72 s.

As a practical example for the assistive technology, we apply the developed interface to the task of humanoid robot control to assist quadriplegic patients, as illustrated in Fig. 1. We first design 12 patient-assisting tasks for the robot to undertake, such as walking, speaking a predefined sentence, and playing music. Second, we implement a two-level hierarchical interface [19] recognizing the tongue and eye movements. The eye movements are used to select the task category, such as walking, dancing, and taking pictures. The tongue movements are used for selecting more specific tasks within the selected category. In addition to the tongue and eye movements, a teeth-clenching detection system using EMG, is also implemented to execute the selected task.

To evaluate the practical usability of the system, we test whether the subjects could control the robot sufficiently to move it to a desired position and then make it perform given tasks. The online experiment results showed that all 12 subjects successfully controlled the robot after practice sessions of less than 30 min.

The rest of this paper is organized as follows. In Section II, we look at the electrophysiology of GKP and EOG. The exemplary signals, their spatial patterns, and the interference between the two signals are also discussed. In Section III, we describe how to separate GKP and EOG signal components, and how to detect the tongue and eye movements. Section IV presents the experimental results from the evaluation of the efficacy and performance of the interface developed. Finally, our results are summarized and discussed in Section V.

II. Background

A. GKP and EOG

GKP refers to the voltage response generated by the tongue movement [15]. The tip of the tongue has a negative charge with respect to the root. As a result, when the tongue touches the inside of a cheek, the electric charge decreases the potential on the skin near the point of contact. In addition, its effect is maximized at the contact point and decreases with distance from the point [20, pp. 294]. Thus, if the negatively charged tongue moves to the left side, the potentials on the electrodes in the left facial region will gradually decrease and the potentials on the electrodes in the right facial region will increase proportionally, and vice versa.

The example signals for the horizontal tongue movements are plotted in Fig. 2. The signals were recorded in the facial regions (see Fig. 3) by using a g.Mobilab device (g.tec, Graz, Austria). The device was originally developed for EEG recording, but all GKP, EOG, and EMG signals are electric potential responses from the facial movements, so the device can be used for recording these signals as well. As shown in the picture located on the

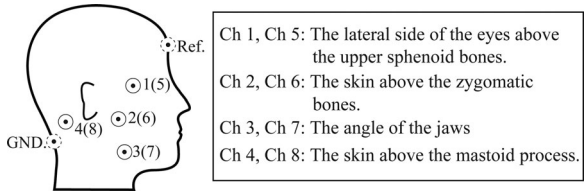

Fig. 3. First four channels (Ch1–Ch4) were attached on the left side of the face and the final four channels (Ch5–Ch8) were attached on the right side.

far left of Fig. 1, the electrodes were attached to the skin by medical tape. The sampling rate was 256 Hz, and the signals were filtered using a high-pass filter above 0.5 Hz.

The first three sessions showing in Fig. 2 labeled by $T_{(1)}$, $T_{(2)}$, and $T_{(3)}$, (where "T" represents "tongue") correspond to the potential responses for the horizontal tongue movements. During the recording, three cues—left (L, touching the inside of the left cheek), front (F, touching the lips to the front), and right (R, touching the inside of the right cheek)—were sequentially shown to the subjects. The two lines labeled "Tongue" and "Eye" represent the cue for each movement. In accordance with the cues, a subject continuously moved his tongue from L to F to R, which is an action akin to licking the lips from the left to the right sides. The results clearly showed that when the tongue moved to the right, the potential levels in the left facial region (Ch1–Ch3) increased marginally, whereas the potential levels in the right facial region (Ch5–Ch7) decreased.

The final three sessions in Fig. 2 correspond to the potential response for the horizontal eye movements. In each session, three auditory cues for left (L), front (F), and right (R) were sequentially given to the subject. The auditory cues were applied to prevent the gaze direction from being distracted by seeing the visual cues. During the experiments, the subject was asked to gaze in the direction of the given cues. The results for $E_{(1)}$, $E_{(2)}$, and $E_{(3)}$ in Fig. 2 (where "E" represents "eye") indicated that when the gaze direction moved to the right, the potentials for the left side (Ch1) decreased, whereas those for the right side (Ch5) increased.

B. Tongue Direction and Gaze Detection

In the previous two sections, we showed that the horizontal tongue and eye movements result in antisymmetric potential changes in the facial region. In other words, the potential differences between the electrodes in the left- and right-side regions can be translated to the horizontal directions of the tongue and gaze. A simple subtraction between the potentials in the left and right channels may be the simplest way to do this, but in order to maximize the sensitivity, we utilized a principal component analysis (PCA)-based feature extraction method.

The eight channel signals $\boldsymbol{X}$ in Fig. 2 can be segmented into the tongue movement-related signals $\boldsymbol{X}_{\mathrm{T}}$ and eye movement-related signals $\boldsymbol{X}_{\mathrm{E}}$. $\boldsymbol{X}_{\mathrm{T}}$ corresponds to the first three sessions of $\boldsymbol{X}$, from 3 to 55 s; $\boldsymbol{X}_{\mathrm{E}}$ corresponds to the final three sessions from 63 to 118 s.

To maximize the feature's sensitivity to the tongue movements, we obtained the feature vector $\boldsymbol{u}_{\mathrm{T}}$ that satisfies the optimization problem

$$\boldsymbol{u}_{\mathrm{T}} = \arg\max_{\|\boldsymbol{u}\|=1}(\boldsymbol{u}^{\top}\boldsymbol{X}_{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}}^{\top}\boldsymbol{u}). \tag{1}$$

The normalized vector $\boldsymbol{u}_{\mathrm{T}}$ with unit length 1 is visualized in Fig. 4(a) by a topographic map (for the convenience, all spatial filters in this paper will be normalized in the same way). $\boldsymbol{u}_{\mathrm{T}}$ gave weights with different signs to the channels of the left and right sides to maximize the feature's sensitivity. The feature values $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ are also obtained and depicted in Fig. 4(c). The results of the first three sessions (shaded region), showed that $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ followed the shape of the cue for the tongue movements, which is presented below.

We also applied the PCA-based feature extraction method to $\boldsymbol{X}_{\mathrm{E}}$. The feature vector $\boldsymbol{u}_{\mathrm{E}}$ that satisfies

$$\boldsymbol{u}_{\mathrm{E}} = \arg\max_{\|\boldsymbol{u}\|=1}(\boldsymbol{u}^{\top}\boldsymbol{X}_{\mathrm{E}}\boldsymbol{X}_{\mathrm{E}}^{\top}\boldsymbol{u}) \tag{2}$$

and the corresponding feature values $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$ are also illustrated in Fig. 4(b) and (c). Unlike $\boldsymbol{u}_{\mathrm{T}}$, the filter $\boldsymbol{u}_{\mathrm{E}}$ imposed larger weights on Ch1 and Ch5, which are located near the eyes. As in the GKP experiment, the feature values $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$ followed the shape of the EOG cue, but in reverse. The results clearly showed that the EOG response can be used to detect the direction of the gaze, as already explained in [4].

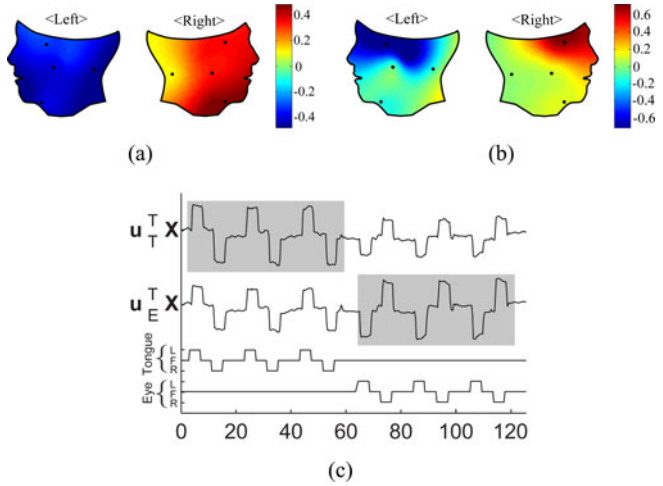

Fig. 4. Feature vectors ($\boldsymbol{u}_{\mathrm{T}}$, $\boldsymbol{u}_{\mathrm{E}}$) and feature values ($\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$, $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$) for detecting the tongue and eye movements. (a) Feature vector $\boldsymbol{u}_{\mathrm{T}}$. (b) Feature vector $\boldsymbol{u}_{\mathrm{E}}$. (c) Feature values $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$.

C. Interference Between GKP and EOG

From the results for $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$, we showed that each $\boldsymbol{u}_{\mathrm{T}}$ and $\boldsymbol{u}_{\mathrm{E}}$ is sensitive to the horizontal tongue and eye movements, and they can be used as feature vectors to detect the directions of the tongue and gaze. However, we also observed that the patterns of the GKP and EOG responses are similar and they can interfere with each other. In the results depicted in Fig. 4(a) and (b), both the EOG and GKP responses exhibited an antisymmetric spatial pattern. Owing to this similarity, $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ was sensitive not only to the tongue movements but also to the eye movements. From the final three sessions in Fig. 4(c), $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ fluctuated considerably according to the eye movements. In the same way, $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$ was also sensitive to the tongue movements, as

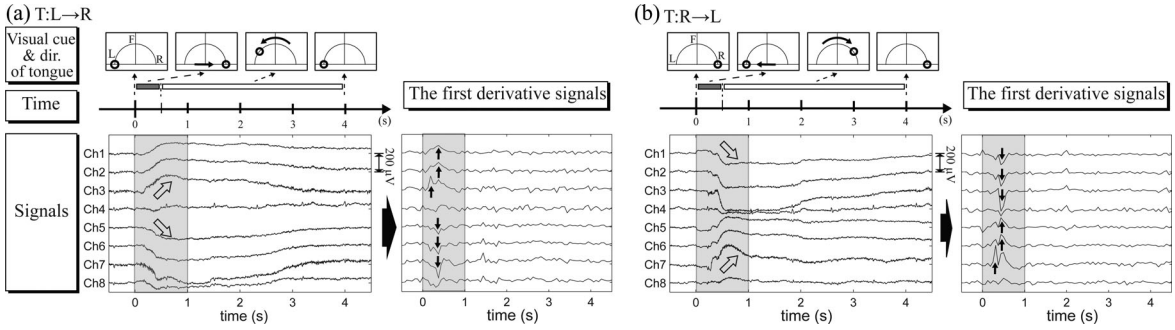

Fig. 5. Cues and exemplary signals for the tongue movements: (a) T:L→R and (b) T:R→L. To record the signals, we showed visual cues to a subject and asked him to move his tongue in the direction of the cue. For the T:L→R movement, a cue in the left side quickly moved to the right side within 0.5 s before slowly returning to the left side for 3.5 s along an upper semicircular path. When the tongue moved to the right side (shaded region), the potential levels on the left side (Ch1–Ch3) increased, whereas the levels on the right side (Ch5–Ch7) decreased (blank arrows). After 1 s, with the tongue moved back to the left side, the potential levels slowly returned to the original levels. The potential changes in 0–1 s generated the spikes denoted by the black arrows in the first derivative signals. However, the slow returning movements in 1–4 s did not generate visible changes in the first derivative signals. For the T:R→L movement, the cue moved in the reverse direction of (a). Potential changes and spikes also appeared at the same time points but in reverse directions.

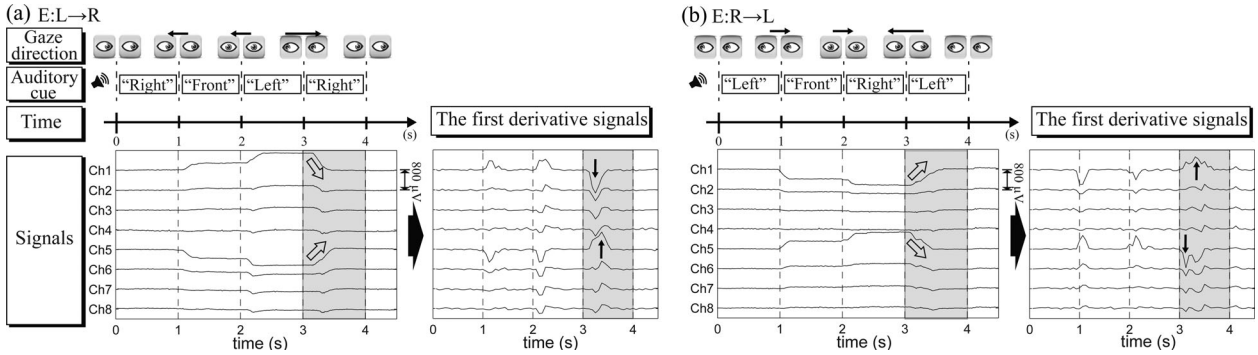

Fig. 6. Cues and exemplary signals for the eye movements: (a) E:L→R and (b) E:R→L. For the movement of E:L→R in (a), the four auditory cues–for "right," "front," "left," and "right"– were sequentially given to the subject every second and we asked him to gaze in the cue's direction. When the subject's gaze moved from left to right in 3–4 s (shaded region), the potential of Ch1 decreased, whereas that of Ch5 increased. These potential changes were also transformed into the spikes of the first derivative signals, as denoted by black arrows. When the gaze moved from right to left in (b), the potential of Ch1 increased whereas that of Ch5 decreased. Subsequently, the corresponding derivative signals showed a positive spike in Ch1 and negative spike in Ch5.

shown in the first three sessions in the figure. This interference was commonly observable from all the subjects participated in the next experiments due to the characteristics of PCA.

III. Method

To solve this interference problem, we designed a new experimental design and an analysis model that can selectively emphasize its sensitivity to the tongue and eye movements. This section explain how we record the training set and build the detection and classification models from the set.

A. Signal Acquisition for Target Movements

Our objective is to develop the interface that can detect the occurrence of the tongue and eye movements and differentiate their directions as accurately as possible. To implement the interface, we selected the following four target movements for detection: 1) T:L→R (tongue movement from left to right), 2) T:R→L, 3) E:L→R (eye movement from left to right), and 4) E:R→L. These movements, which move from the lateral sides to the opposite side by 180° (instead of the movement from the medial side by 90°), were selected to reject the false positives from task-irrelevant movements. The movements from the medial side frequently occur involuntarily in daily life, especially in the case of eye movement. Therefore, they are not suitable for use in the interface. Moreover, the movements between the lateral sides generate larger potential changes than the movements between the medial and lateral sides are therefore easier to detect.

To train the detection model, we recorded the signals over seven rounds of tongue and eye movements. The given cues and recorded signals are illustrated in Figs. 5 and 6. To record the signals for the tongue movements in Fig. 5, we showed visual cues to the subject and asked him to move his tongue in the direction where the cue was located, while maintaining contact between the tongue and the inside of the cheek. To record the signals for the eye movements in Fig. 6, we gave the subject four auditory cues assigned to three directions and then asked the subject to gaze in the cue's direction. As explained in the previous section, the potential levels changed in an antisymmetric way with the tongue and eye movements. In particular, potential changes in the tongue movements were commonly observed in Ch1−Ch3 and Ch5−Ch7, whereas changes in the eye movements were dominant in Ch1 and Ch5, as indicated by blank arrows. The signals for movements to different directions resulted in inverse waveforms.

For each kind of movements, the cues were given seven times. The length of each trial was 4 s, and the trials were interspersed with 3 s of intertrial blanks. In total, recording a single session for 28 trials took 196 s.

B. Using the First Derivative Signals

In the next step, we measured the mean values of the recorded potentials from each channel for every 0.06 s (the duration was empirically decided based on the preliminary tests) and then obtained the first derivative of the mean values, $\boldsymbol{x}'(t) = \boldsymbol{x}(t) - \boldsymbol{x}(t-1)$. The first derivative signals were measured to eliminate the effect of drift and precisely detect the time of movements. When recording the bioelectrical potential, the baseline of the signal can shift across time owing to the drift [21, pp. 25–27]. However, the first derivative signals are robust against this artifact because its effect cannot accumulate. Furthermore, the response for the movements is easier to detect on the first derivative signals than on the raw signals. As shown in Figs. 5 and 6, the potential changes evoked by the sudden movements were transformed to positive and negative spikes, as denoted by the black arrows. The temporal locations of these sharp spikes were easier to specify than the potential changes in the raw signals.

In detail, in the first derivative signals for the T:L→R movement, Ch1–Ch3 gave positive spikes while Ch5–Ch7 gave negative spikes. In contrast, the signals for the T:R→L movement gave negative spikes on Ch1–Ch3, and positive spikes on Ch5–Ch7. The signals for E:L→R and E:R→L also gave similar spikes but in the different channels. The signals for E:L→R gave a negative spike in Ch1 and a positive spike in Ch5, while E:R→L gave inverse spikes in the same channels. Therefore, by checking the directions of the spikes and the channels where they are located, we can distinguish the four movements.

C. Discriminative Feature Extraction

From these first derivative signals, we obtained a 2-D feature for detection. Let us redefine $\boldsymbol{X}$ as the first derivative signal. From $\boldsymbol{X}$, we collected the signals of the first 14 trials as $\boldsymbol{X}_{\mathrm{T}}$, which correspond to the tongue movements. The signals for the last 14 trials were collected as $\boldsymbol{X}_{\mathrm{E}}$, which correspond to the eye movements. For the initial approach to feature extraction, we obtained $\boldsymbol{u}_{\mathrm{T}}$ and $\boldsymbol{u}_{\mathrm{E}}$ by (1) and (2), respectively. Fig. 7 shows the projection results for $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$. As already noted in the previous section, both features were sensitive to the tongue and eye movements.

In order to solve the interference problem, we sought a linear projection discriminating the tongue movement from the eye movement. We considered the *generalized Rayleigh quotient*, which is defined as

$$\mathcal{R}(\boldsymbol{X}_{\mathrm{T}}, \boldsymbol{X}_{\mathrm{E}}, \boldsymbol{w}) = \frac{\boldsymbol{w}^{\top}\boldsymbol{X}_{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}}^{\top}\boldsymbol{w}}{\boldsymbol{w}^{\top}\boldsymbol{X}_{\mathrm{E}}\boldsymbol{X}_{\mathrm{E}}^{\top}\boldsymbol{w}} \tag{3}$$

and represents the ratio of projected variances of tongue-only data $\boldsymbol{X}_{\mathrm{T}}$ and eye-only data $\boldsymbol{X}_{\mathrm{E}}$. The generalized Rayleigh quotient reaches its maximum value when the tongue's variation is maximized while the eye's variation is minimized. Therefore,

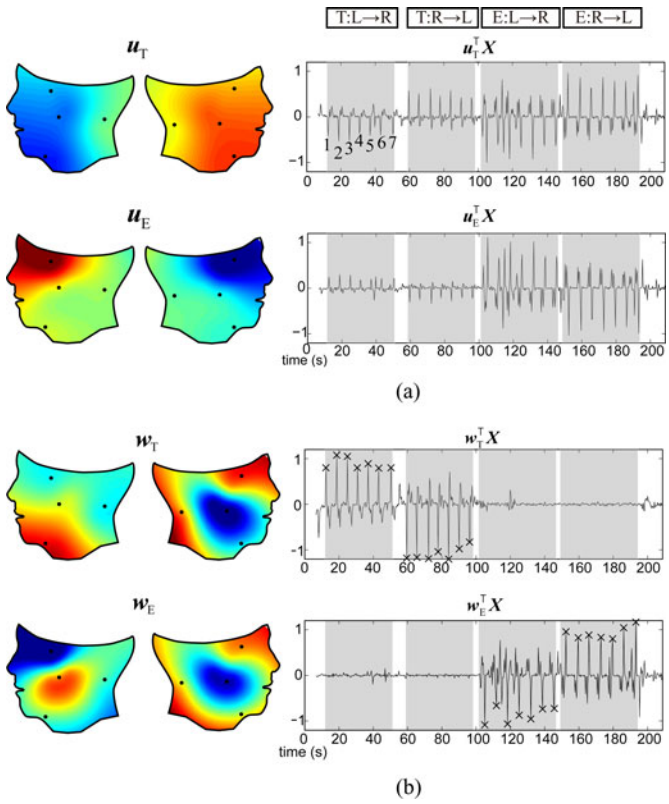

Fig. 7. Feature values for the tongue and eye movements obtained by the PCA and the discriminative feature extraction method. For each group of movements, the feature values in (a) showed seven spikes, each of which corresponds to a single trial of the movements. $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$, obtained by PCA, responded to both the tongue and eye movements. In the case of the discriminative method, $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ responded only to the tongue movements and $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$ responded only to the eye movements. In $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$, the spikes in each group were clearly distinguishable from each other by their direction and height. To build the detection model, we collected a feature point of the positive class from each trial, as denoted by "×" marks. (a) Features from PCA. (b) Features from discriminative feature extraction.

we determined $\boldsymbol{w}_{\mathrm{T}}$ and $\boldsymbol{w}_{\mathrm{E}}$ such that

$$\boldsymbol{w}_{\mathrm{T}} = \arg\max_{\|\boldsymbol{w}\|=1} \mathcal{R}(\boldsymbol{X}_{\mathrm{T}}, \boldsymbol{X}_{\mathrm{E}}, \boldsymbol{w}) \tag{4}$$

$$\boldsymbol{w}_{\mathrm{E}} = \arg\max_{\|\boldsymbol{w}\|=1} \mathcal{R}(\boldsymbol{X}_{\mathrm{E}}, \boldsymbol{X}_{\mathrm{T}}, \boldsymbol{w}). \tag{5}$$

Since $\boldsymbol{X}_{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}}^{\top}$ and $\boldsymbol{X}_{\mathrm{E}}\boldsymbol{X}_{\mathrm{E}}^{\top}$ form a symmetric-definite matrix pencil, the maximum or minimum of $\mathcal{R}(\boldsymbol{X}_{\mathrm{T}}, \boldsymbol{X}_{\mathrm{E}}, \boldsymbol{w})$ can be easily determined by solving the *generalized eigenvalue problem*

$$(\boldsymbol{X}_{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}}^{\top})\boldsymbol{w} = \lambda(\boldsymbol{X}_{\mathrm{E}}\boldsymbol{X}_{\mathrm{E}}^{\top})\boldsymbol{w} \tag{6}$$

which also satisfies

$$\begin{cases} \boldsymbol{W}^{T}\boldsymbol{X}_{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}}^{\top}\boldsymbol{W} = \boldsymbol{\Lambda}_{\mathrm{T}}, \\ \boldsymbol{W}^{T}\boldsymbol{X}_{\mathrm{E}}\boldsymbol{X}_{\mathrm{E}}^{\top}\boldsymbol{W} = \boldsymbol{\Lambda}_{\mathrm{E}} \end{cases} \tag{7}$$

where $\boldsymbol{\Lambda}_{\mathrm{T}}$ and $\boldsymbol{\Lambda}_{\mathrm{E}}$ are diagonal matrices [22, pp. 31–34]. To specify the scale of $\boldsymbol{W}$, let us put $\boldsymbol{\Lambda}_{\mathrm{E}}$ to $\boldsymbol{I}$, the identity matrix. Then, we can measure $\boldsymbol{w}_{\mathrm{T}}$ and $\boldsymbol{w}_{\mathrm{T}}$ by calculating the eigenvectors associated with the largest or smallest eigenvalue of $(\boldsymbol{X}_{\mathrm{E}}\boldsymbol{X}_{\mathrm{E}}^{\top})^{-1}(\boldsymbol{X}_{\mathrm{T}}\boldsymbol{X}_{\mathrm{T}}^{\top})$. The obtained spatial filters ($\boldsymbol{w}_{\mathrm{T}}$ and

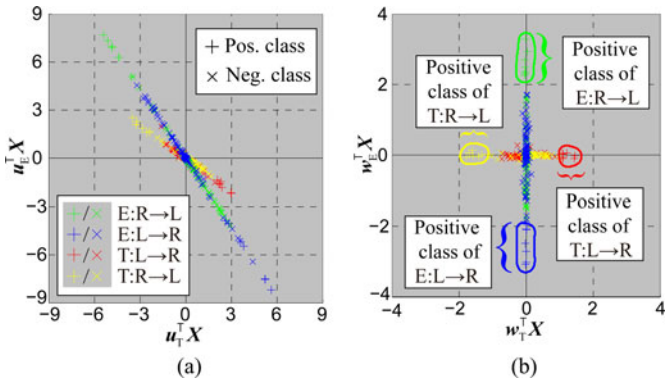

Fig. 8. Feature space and decision boundaries for detection. We plotted the feature value pairs $(\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}, \boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X})$ in (a) and $(\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}, \boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X})$ in (b). As shown in (b), the points in each positive class of the four movements were in clear clusters distinguishable from the points in the negative class. (a) PCA. (b) Discriminative method.

$\boldsymbol{w}_{\mathrm{E}}$) and their projection results ($\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$) are plotted in Fig. 7.

In contrast to the vectors for $\boldsymbol{u}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{u}_{\mathrm{E}}^{\top}\boldsymbol{X}$ from the PCA, the vectors for $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$ could selectively emphasize the tongue and eye movements. In the first 14 trials of the signals, which corresponded to the tongue movements, only $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ fluctuated significantly. In the last 14 trials for the eye movements, only $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$ fluctuated and $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ remained stable.

To verify the usability of the discriminative feature extraction method, we plotted the feature space obtained from PCA in Fig. 8(a), as a counter example. Due to the similarity between GKP and EOG, there was a correlation between the feature points for the eye movements (green and blue dots) and the tongue movements (red and yellow dots). This correlation fails to make a clear decision for the classification. However, with the discriminative feature extraction method, this correlation was removed, as shown in Fig. 8(b). In the figure, the horizontal axis was sensitive only to the tongue movements and the vertical axis was sensitive only to the eye movements.

This orthogonalized feature space also provides intuitive visual feedback, which helps the subject to easily adapt to the interface. During the testing procedure, we showed the feature space and current feature point to the subjects in real time. This visual feedback provide the subjects useful information, such as the sensitivity, response time of the interface, and the optimal speed of the tongue and eye movements, which is essential for adapting to the interface.

D. Classification

In the previous section, we showed that $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ and $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$ are useful features for discriminating the tongue and eye movements and differentiating their directions. However, the objective of this paper is to not only differentiate the movements but also detect their occurrences precisely. For this purpose, we separated the obtained feature points into positive and negative classes and then trained the model to detect this positive class in the testing procedure. The positive class corresponds to the signals recorded at the exact moment when the target movement occurred, whereas the negative class corresponds to the signals recorded when the tongue and eyes returned slowly to the original positions. This negative class can represent slow and unintended movements that can cause false positives. The feature extraction method transforms each target movement into a single spike in $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ or $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$. We collected the points located at these spikes as the positive class. We selected the point with the maximum or minimum feature value from each trial and then added it to the positive class of the trial's movement, while all other points were added to the negative class. In Fig. 7(b), the temporal locations of the points in the positive classes are denoted by "×" marks. In Fig. 8, the points in the positive classes of each movement are denoted by the "+" marks, whereas those in the negative classes are denoted by "×" marks. In Fig. 8(b), the points in the positive classes have four clear clusters that are distinguishable from the points in the negative classes.

From the feature space, we obtained the decision boundaries to detect four kinds of tongue and eye movements. For each movement, we set the feature points from the positive class of the movement of interest as one class and the feature points from the negative classes of all movements as another class. Then, we trained a radial basis function kernel support vector machine (RBF SVM)-based classifier for each movement by using the LIBSVM library [26]. The decision boundaries obtained from four binary classifiers are drawn in the Fig. 8(b). Despite its complexity compared to a linear SVM, we applied RBF SVM, because its enclosed decision boundary is useful for rejecting the motion artifacts that arise when the electrodes tremble or are moved. Generally, this artifact arises when the subjects move their eyes or tongue too fast or when they move their head. In this case, the potential levels drastically fluctuate, and the corresponding feature points radically diverge. The enclosed decision boundary of RBF SVM can reject these outliers and reduce false positive detections.

E. Humanoid Robot Control

The objective of our interface is to assist patients by enabling them to execute various robot tasks, such as delivering a message to someone or providing entertainment to the patients themselves. For this purpose, we implemented a two-level hierarchical interface [19] that can instruct a robot to perform 12 kinds of tasks in six categories by the tongue and eye movements. We used the NAO robot system from Aldebaran Robotics [see Fig. 9(a)]. This is a 57-cm tall humanoid robot that has a body with 25 movable joints. Its legs enable it to walk in any direction. In addition to the walking ability, the robot also has a speaker to play music, a camera to take pictures, and a voice synthesizer that can read various English words.

Based on these various abilities, we implemented a robot control interface system that performs the various tasks depicted in Fig. 9(c). The state diagram in the figure also explains how the tasks were changed and executed. First, we enabled six categories of tasks for the robot system as represented by the rounded boxes. Each category has specific tasks that are represented by circles. The eye-movement detection system was used for category selection, and the tongue movement detection system was used for specific task selection. We also implemented an EMG-based teeth-clenching detection system when task approval is

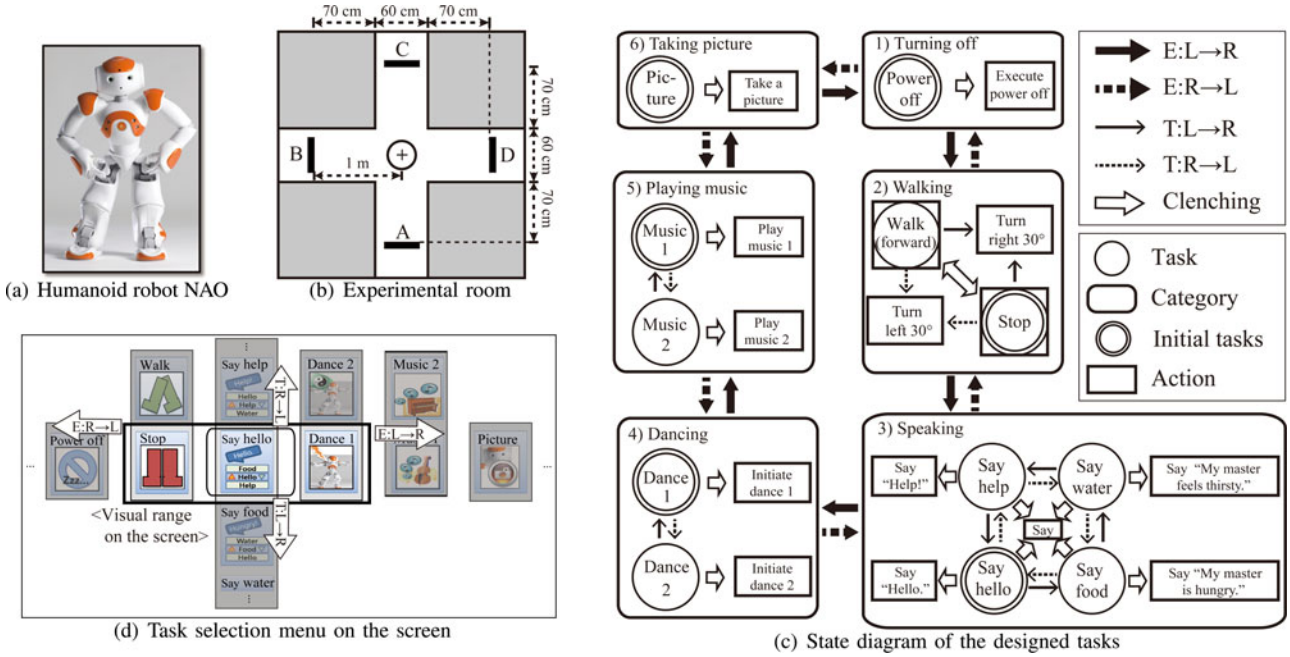

Fig. 9. Developed robot control interface and experimental environment. The developed interface was applied to the control of the humanoid robot NAO in (a). During the experiments, we asked the subjects to navigate the robot in the room [depicted in (b)] from one position to another and then perform one of the tasks in (c). The figures in (c) and (d) explain how the subjects changed and executed the tasks. As shown in (c), the categories could be selected by the eye movements. That is, if an E:L→R movement was detected, the focus for the category moved to the clockwise direction and vice versa, as represented by the thick arrows. Within the categories, a specific task was selected by the tongue movements and executed by a clenching movement. Particularly, in the "walking" category, the robot was advanced by toggling a teeth-clenching movement to switch the task state between "walk" and "stop." The figure in (d) shows the menu represented on the screen to enable the subjects to control the interface as depicted in (c). On the screen, three tasks were represented in the visual range. The center position showed the current task, which could be executed by a teeth-clenching movement. This visual range was shifted in the horizontal direction by E:R→L and E:L→R movements, and the task in the center position was shifted in vertical directions by T:R→L and T:L→R movements.

required. To detect the clenching movements, the variance of all channels was measured and then compared to the threshold, as already explained in [1].

IV. Results

To evaluate the performance of our system and prove its usability, we designed three types of experiments. In the first experiment (performance evaluation experiment), we evaluated the accuracy and response time of the interface. In the second online experiment, we asked our subjects to control the robot, move it to a desired position, and making it perform tasks such as talking, playing music, and taking pictures. In the third experiment, we surveyed the subjective workload with NASA-Task Load indeX (TLX) test to evaluate the performance in qualitative measures from the viewpoint of the subjects. For the experiments, we recorded a single training dataset from each subject, as described in Section III-A. The results for a sevenfold cross-validation test are noted in Table I. The analysis model obtained from the dataset was repeatedly used for all the three types of experiments.

TABLE I
Detection Accuracies (%) and Response Time (s, Inside the Bracket)

Sub.	1	2	3	4	5	6
CV	89.3	92.9	82.1	92.9	82.1	85.7
S. 1	95 (3.28)	90 (3.96)	80 (3.77)	80 (3.45)	75 (2.59)	80 (1.93)
S. 2	95 (3.97)	85 (3.80)	85 (3.06)	85 (2.93)	90 (2.58)	85 (2.87)
S. 3	90 (3.18)	90 (3.78)	95 (3.05)	80 (3.24)	85 (2.09)	80 (2.27)
Mean	93.3 (3.48)	88.3 (3.85)	86.7 (3.29)	81.7 (3.21)	83.3 (2.42)	81.7 (2.36)
Sub.	7	8	9	10	11	12
CV	89.3	82.1	100.0	71.4	89.3	92.9
S. 1	90 (2.21)	70 (1.96)	100 (3.11)	75 (2.82)	95 (2.02)	80 (3.74)
S. 2	95 (2.25)	80 (1.65)	90 (3.08)	85 (1.88)	100 (1.92)	90 (3.32)
S. 3	75 (2.11)	100 (1.63)	100 (3.00)	75 (2.09)	80 (1.93)	95 (3.29)
Mean	86.7 (2.19)	83.3 (1.75)	96.7 (3.07)	78.3 (2.26)	91.7 (1.96)	88.3 (3.45)

Sub.: Subject, CV: Cross-Validation, S.: Session

A. Performance Evaluation Experiment

In the first experiment, we evaluated the accuracy and response time of the interface. As explained in Section III, the interface could detect four kinds of movements. We gave 20 random cues belonging to one of the four movements to the subjects, and asked them to perform the movement of the given cue. We then checked to ascertain whether the system identified the movements correctly, and we measured the response times.

Eleven male and one female subjects participated in the experiment, and they aged 23–31 years (mean = 27.3). Subject 1 was familiar in the use of the interface, whereas the others were naive subjects who were using the interface for the very first time. The experimental results are presented in Table I. The mean accuracy was 86.7 $\pm$ 8.28%, whereas the mean response time was 2.77 $\pm$ 0.72 s. The average accuracy of EOG trials was 92.6% and the accuracy of GKP trials was 78.7%. Compared to

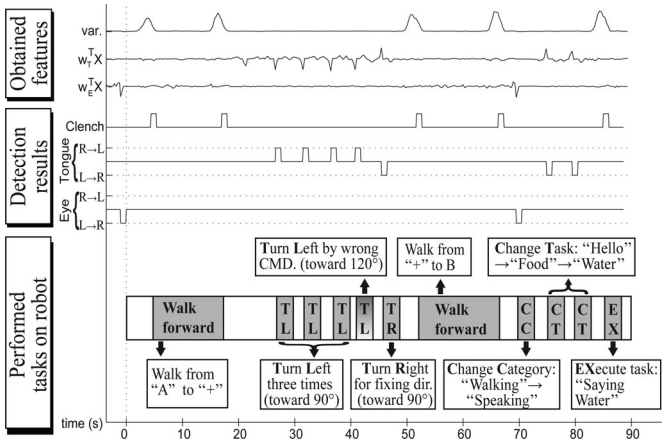

Fig. 10. Results of the experiment for subject 1, session 1. Before the experiment, the "power off" task was selected as the initial task. The experiment began when E:L→R was detected at 0 s, which changed the selected task to "stop" in the "walking" category. When the clenching movement was detected at 4.4 s, the task was changed to "walk," and the robot began to walk forward from A. When the robot reached to "+," the subject clenched his teeth once again to stop the robot. To steer the robot to the left toward B, the subject was required to repeat T:R→L movement three times to turn 90°. However, during the experiment, he moved it four times erroneously (the rotation speed of the robot was slower than he expected, so he thought he should repeat it four times). To rectify this additional turning, he moved his tongue in the reverse direction (T:L→R) once at 45.5 s. He repeated two clenching actions, at 51.7 and 66.3 s, to move the robot from "+" to B. After the robot arrived at B, the subject initiated movements to execute the action cue. He performed E:L→R movement at 69.5 s to change the category from "walking" to "speaking." Then, he repeated T:L→R movement twice to change the task from "hello" to "food" to "water." After that, he clenched his teeth to execute the currently selected "water" task; the robot then read the predefined sentence.

the recent EOG classification research in [13], which reported 95% of accuracy, our system showed slightly lower accuracy.

B. Online Experiment

In the second experiment, we tested whether the subjects could control the robot sufficiently to navigate it to a specific position in the room and then have it perform the desired tasks. The layout of the room used in this experiment shown in Fig. 9(b). "A," "B," "C," and "D" represent four destination positions, and the gray regions in the figure represent walls (constructed using carton boxes). Before each experiment, the robot was set at one of the four positions. When the experiments began, we gave the cues in pairs consisting of a "destination cue" and an "action cue." The destination cue specified one of the destination positions (A, B, C, or D), whereas the action cue specified one of the tasks in Fig. 9(c). We asked the subjects to control the robot with the developed interface, make it reach the position where the destination cue was located, and then have the robot perform the task in the action cue. The video clip for the exemplary experiment is available at http://mlg.postech.ac.kr/research/GOMFace.html.

We performed the experiments with 12 subjects who participated in the performance evaluation experiments. The example result of subject 1 is plotted in Fig. 10. The figure shows the obtained features, detection results, and corresponding robot actions. The first feature value (sum of variances from all channels) increased when the subjects clenched their teeth. The second feature $\boldsymbol{w}_{\mathrm{T}}^{\top}\boldsymbol{X}$ increased with T:L→R movement, and decreased with T:R→L movement. The third feature $\boldsymbol{w}_{\mathrm{E}}^{\top}\boldsymbol{X}$ increased with E:R→L movement and decreased with E:L→R. From these feature values, five kinds of movements were detected as shown in the middle of the figure. Fig. 10 also illustrates how the detection results commanded the robot to perform actions. As shown in the figure, even if the subject entered the wrong command as input, he could fix the result by entering an additional error correction command.

The experimental results for all subjects are listed in Table II. The taken time (t_{T}) indicates the time record of the experiments to fulfill the destination and action cues. In the next column ("# of CMDs"), we denoted the number of commands in ideal case (n_{I}), which is the minimum number of movements to complete the task. For example, if "A to B" and "say water" cues are given as shown in Fig. 10, 11 movements (seven for navigating the robot from A to B and four for selecting and executing the "say water" task) should be correctly detected by the system. In the same column, we also denoted the number of movements actually input by the subjects inside the bracket. The difference between the two values indicates the number of mistakenly detected movements or additional movements to correct them.

From t_{T}, we measured the average response time (t_{A}) required to input a single valid command. t_{T} consists of the time for robot movements (t_{R}) and the time for interface manipulation (t_{I}). Whenever a command was sent to the robot, the subjects should wait until the robot finished walking or tuning movements, and t_{R} is corresponding to these waiting times. Our concern is how much time did the subjects spend to control the interface ($=t_{\mathrm{I}}$), so t_{R} should be removed from the performance evaluation. For all the experiments in Table II, t_{R} had the same value, because for all destination cues, the robot was supposed to walk 1 m, turn 90°, then walk 1 m again. To obtain t_{R}, we measured the average elapsed time for the robot to perform these three movements; the measured t_{R} was 46.3 ± 1.49 s. We measured the time for interface manipulation by $t_{\mathrm{T}} - t_{\mathrm{R}}$, then, calculated the average response time t_{A} by dividing this value by n_{I}, as $t_{\mathrm{A}} = (t_{\mathrm{T}} - t_{\mathrm{R}})/n_{\mathrm{I}}$.

The total average of all subjects' response time was 4.44 ± 1.46 s. In general, the records for the online experiments were longer than those for the offline experiments. During the online experiments, there is a possibility of wrong commands being input; in this case, the subjects should input additional commands to rectify them. In addition, if the system fails to recognize the movement, the subjects should repeat the movement until it is correctly recognized. These malfunctions increased the response time of the experiments from 2.77 to 4.44 s.

C. Workload Comparison Experiment

In addition to the previous experiments, which evaluated the performance in quantitative measures by accuracy and response time, in the third experiment, we evaluated the performance in qualitative measure by surveying the subjective workload with NASA-TLX. Hart and Staveland broke workload into six different factors (mental demand, physical demand, temporal demand, effort, performance, and frustration) [24], and NASA developed the NASA-TLX, which is a measurement tool for estimating these factors [25]. By using this tool, we evaluated

TABLE II
TIME TAKEN (S) AND AVERAGE RESPONSE TIME (S) FOR THE ONLINE EXPERIMENTS

Subject	Dest. cue	Action cue	Taken time (t_T)	# of CMDs*	Avr. response time (t_A)	Dest. cue	Action cue	Taken time (t_T)	# of CMDs*	Avr. response time (t_A)
1	A to B	Say water	85.1	11 (13)	3.53	B to C	Dance 2	79.4	11 (12)	3.01
	D to A	Take picture	84.2	10 (13)	3.79	C to B	Music 1	75.2	11 (11)	2.63
2	A to D	Take picture	100.0	10 (14)	5.37	A to B	Say hello	83.6	9 (12)	4.14
	B to C	Dance 2	82.6	11 (14)	3.30	A to B	Take picture	94.4	10 (17)	4.81
3	A to D	Take picture	88.0	10 (13)	4.17	C to D	Music 1	70.8	11 (11)	3.01
	D to A	Say food	97.0	10 (14)	5.07	C to B	Dance 1	86.8	10 (11)	4.05
4	A to D	Take picture	126.4	10 (16)	8.01	A to D	Music 1	70.8	11 (13)	3.72
	B to C	Dance 2	87.2	11 (15)	3.72	B to A	Say hello	68.2	9 (9)	2.43
5	A to B	Dance 1	109.4	10 (12)	6.31	B to C	Say help	82.4	10 (11)	3.61
	C to D	Take picture	94.4	10 (12)	4.81	D to A	Music 1	122.6	11 (14)	6.94
6	D to A	Take picture	119.2	10 (15)	7.29	A to D	Say water	87.3	11 (11)	3.72
	D to C	Music 2	97.9	12 (12)	4.69	C to B	Say help	94.0	10 (12)	4.77
7	A to D	Say help	89.7	10 (12)	4.34	A to D	Dance 2	105.6	11 (13)	5.39
	B to C	Music 1	122.3	11 (17)	6.90	B to A	Say food	94.1	10 (12)	4.78
8	A to B	Dance 1	78.7	10 (12)	3.24	C to D	Music 1	83.9	11 (14)	3.41
	B to C	Dance 1	89.4	10 (12)	4.11	D to A	Say help	77.0	10 (11)	3.07
9	A to D	Dance 1	80.9	10 (12)	3.46	D to C	Say water	79.5	11 (14)	3.02
	C to D	Say food	71.3	10 (12)	2.50	D to A	Music 2	83.5	12 (11)	3.10
10	A to B	Say help	112.6	10 (16)	6.63	B to C	Dance 1	92.0	10 (12)	4.57
	C to D	Music 2	82.9	12 (12)	3.05	D to A	Dance 2	80.4	11 (13)	3.10
11	A to B	Dance 2	110.3	11 (12)	5.82	B to A	Say food	110.0	10 (14)	6.37
	A to D	Take picture	83.0	10 (12)	3.67	D to C	Music 1	94.9	11 (11)	4.42
12	A to D	Say water	92.0	11 (11)	4.15	D to A	Music 1	138.7	11 (16)	8.40
	A to B	Take picture	91.6	10 (10)	4.53	B to A	Dance 2	92.3	11 (11)	4.19

*The number of movements actually input by the subjects is denoted inside the bracket.

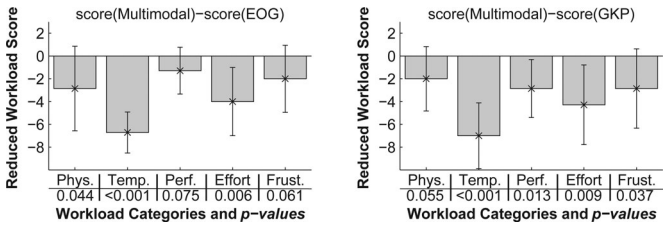

Fig. 11. Workload reduction achieved by multimodal approach.

the workload reduction achieved by multimodal approach. For this experiment, we implemented two unimodal interfaces: the EOG-only and GKP-only interfaces. In these interfaces, we relocated the tasks in Fig. 9(c) with serializing them in a single line. In the EOG-only interface, the subjects could use only the eye movements to change tasks, while in the GKP-only interface, they could use only the tongue movements. In the original multimodal interface, only six eye movements were required to go around all categories and return to the original task. However, in the GKP- and EOG-only interfaces, 13 movements were required for the same job, because category shift function was disabled.

Among 12 subjects, eight subjects participated in the experiment. After finishing one session of online experiment with EOG-only, GKP-only, and original multimodal interfaces, they filled in workload charts with a score range of 0–20. The workload reduction achieved by multimodal approach was calculated by subtracting the workload score of multimodal interface from the scores of EOG- and GKP-only interfaces, and the results are plotted in Fig. 11 with p-values measured by t-test. The score of mental demand was omitted, because no subjects reported the difference in the scores of the three experiments. As shown in the figure, the workloads for the temporal demand were clearly decreased (p-value < 0.001 for both cases) and the workloads for the effort were also decreased (p-value $= 0.006$ and 0.009). As shown in the results, the multimodal interface could relieve the temporal demand and user effort, by enabling the subjects to access the desired task much faster with less effort.

V. CONCLUSION

In this paper, we presented a new human–computer interface system that integrates a GKP-based tongue–machine interface, an EOG-based gaze-tracking interface, and an EMG-based teeth-clenching detection interface to improve the information capacity. In particular, to combat the interference between GKP and EOG, we applied the discriminative feature extraction method for separating the two signal components. The separated signals were classified by using RBF–SVM, and the results were transmitted to a humanoid robot so that it could correctly perform various tasks to assist paralyzed patients. As evidenced by the online experimental results, the developed hierarchical interface system enabled the subjects to select and execute various tasks easily and intuitively. We believe that this integrative approach is an important issue, especially for the quadriplegic patients who have rather limited ways of communicating with machines.

In this paper, we considered only the simple movements in the horizontal directions, but we believe that the further investigations for detecting more complex movements, such as the horizontal eye movements discussed in [13] and the simultaneous tongue and eye movements, are worthwhile. We also expect that the developed system can be integrated with existing EEG-based BCI systems to exploit all recordable electrical potentials from the head region including the face and scalp. We already used the EEG device for recording the electrical potential on the face, so the integration might be easily achieved by adding more electrodes on the scalp. The GKP and EOG systems use a

low-frequency band for analysis, so the interference with ordinary EEG-based BCI systems such as motor imagery and P300 could be negligible.

Moreover, recent BCI can recognize not only the user intention, but also the mental and emotional states of users. For example, Zander [26] detected the occurrence of the loss of control from the EEG signals, and exploited it to improve the robustness of motor imagery BCI system. With these kinds of novel BCI techniques, multimodal interface approach may improve not only the capacity of the interface, but also its robustness and usability.

REFERENCES

[1] Y. Nam, Q. Zhao, A. Cichocki, and S. Choi, "Tongue-Rudder: A glossokinetic potential-based tongue-machine interface," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 290–299, Jan. 2012.

[2] G. Krishnamurthy and M. Ghovanloo, "Tongue drive: A tongue operated magnetic sensor based wireless assistive technology for people with severe disabilities," in *Proc. IEEE Int. Symp. Circuits Syst.*, Island of Kos, Greece, 2006, pp. 5551–5554.

[3] A. Bulling, D. Roggen, and G. Tröster, "Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments," *J. Ambient Intell. Smart Envir.*, vol. 1, no. 2, pp. 157–171, 2009.

[4] T. Yagi, Y. Kuno, K. Koga, and T. Mukai, "Drifting and blinking compensation in electro-oculography (EOG) eye-gaze interface," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Taipei, Taiwan, 2006, pp. 3222–3226.

[5] R. Barea, L. Boquete, M. Mazo, and E. López, "System for assisted mobility using eye movements based on electrooculography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no. 4, pp. 209–218, Dec. 2002.

[6] C. C. Postelnicu, F. Girbacia, and D. Talaba, "EOG-based visual navigation interface development," *Expert Syst. Appl.*, vol. 39, pp. 10 857–10 866, 2012.

[7] W. Nutt, C. Arlanch, S. Nigg, and G. Staufert, "Tongue-mouse for quadriplegics," *J. Micromech. Microeng.*, vol. 8, pp. 155–157, 1998.

[8] X. Huo and M. Ghovanloo, "Using unconstrained tongue motion as an alternative control mechanism for wheeled mobility," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 6, pp. 1719–1726, Jun. 2009.

[9] S. Oviatt, "Multimodal interfaces," in *The Human-Computer Interaction Handbook*, J. J. A. and S. Andrew, Eds. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2003, pp. 286–304.

[10] R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal human-computer interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, May 1998.

[11] G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, and N. Birbaumer, "The hybrid BCI," *Front. Neurosci.*, vol. 4, pp. 1–11, 2010.

[12] C. S. L. Tsui, P. Jia, J. Q. Gan, H. Hu, and K. Yuan, "EMG-based hands-free wheelchair control with EOG attention shift detection," in *Robotics and Biomimetics*. Piscataway, NJ, USA: IEEE, 2007, pp. 1266–1271.

[13] A. B. Usakli, S. Gurkan, F. Aloise, G. Vecchiato, and F. Babiloni, "A hybrid platform based on EOG and EEG signals to restore communication for patients afflicted with progressive motor neuron diseases," in *Engineering in Medicine and Biology Society*. Piscataway, NJ, USA: IEEE, 2009, pp. 543–546.

[14] B. J. Fisch, *Fisch and Spehlmann's EEG Primer*, 3rd ed. New York, NY, USA: Elsevier, 1999.

[15] D. Klass and R. G. Bickford, "Glossokinetic potentials appearing in the electroencephalogram," *EEG Clin. Neurophysiol.*, vol. 12, p. 239, 1960.

[16] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components," *EEG Clin. Neurophysiol.*, vol. 79, pp. 440–447, 1991.

[17] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[18] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[19] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch, "Earpod: Eyes-free menu selection using touch input and reactive audio feedback," in *Proc. ACM Conf. Human Fact. Comput. Syst.*, 2007, pp. 1395–1404.

[20] F. S. Tyner, J. R. Knott, and W. B. Mayer, *Fundamentals of EEG Technology: Vol. 1: Basic Concepts and Methods*. Baltimore, MD, USA: Lippincott Williams & Wilkins, 1983.

[21] P. Tallgren, "DC–EEG for routine clinical use: Methods and clinical impact," Ph.D. dissertation, Dept. Electr. Commun. Eng., Helsinki Univ. Technol., Espoo, Finland, Dec. 2006.

[22] K. Fukunaga, *An Introduction to Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Academic, 1990.

[23] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, 2011.

[24] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 139–183.

[25] S. A. Brewster and M. G. Crease, "Correcting menu usability problems with sound," *Behav. Inf. Technol.*, vol. 18, no. 3, pp. 165–177, 1999.

[26] T. O. Zander and S. Jatzev, "Context-aware brain-computer interfaces: Exploring the information space of user, technical system and environment," *J. Neural Eng.*, vol. 9, no. 1, p. 016003, 2012.

Yunjun Nam received the B.Sc. degree in electrical and computer engineering from Hanyang University, Seoul, Korea, in 2008. He is currently working toward the Ph.D. degree in the School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang, Korea.

His research interests include brain–computer interface and electroencephalography signal processing.

Bonkon Koo received the B.Sc. degree in biomedical engineering and electrical engineering from Kyunghee University, Seoul, Korea, in 2010. He is currently working toward the Ph.D. degree in the School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang, Korea.

His research interests include human–computer interface and electroencephalography signal processing.

Andrzej Cichocki (M'95) received the Ph.D. and Dr.Sc. (Habilitation) degrees, in electrical engineering from the Warsaw University of Technology, Warsaw, Poland.

He is currently the Senior Team Leader and the Head of the Laboratory for Advanced Brain Signal Processing, at the RIKEN Brain Science Institute, Wakō, Japan. He is coauthor of more than 250 technical papers and four monographs (two of them translated to Chinese).

Dr. Cichocki has been invited to the Council of Canadian Academies Survey of Science and Technology Strengths as an author and coauthor of one of the top 1% most highly cited papers in his field worldwide.

Seungjin Choi (M'93) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1987 and 1989, respectively, and the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1996.

He was with the Laboratory for Artificial Brain Systems, RIKEN, Japan, in 1997, and was an Assistant Professor in the School of Electrical and Electronics Engineering, Chungbuk National University from 1997 to 2000. Since 2001, he has been a Professor of Computer Science at Pohang University of Science and Technology, Pohang, Korea. His primary research interests include machine learning and probabilistic models with their applications to data mining, computer vision, and brain–computer interface.