Optimizing interoperability between video-oculographic and electromyographic systems

Javier Navallas, PhD;[1]* Mikel Ariz, MSc;[1] Arantxa Villanueva, PhD;[1] Javier San Agustín, PhD;[2] Rafael Cabeza, PhD[1]
[1]*Department of Electrical and Electronic Engineering, Public University of Navarra, Campus de Arrosadia, Pamplona, Navarra, Spain;* [2]*Innovative Communication Group, IT University of Copenhagen, Copenhagen, Denmark*

Abstract—A new system is presented that enhances the interoperability between a video-oculographic (VOG) system for mouse movement control and an electromyographic (EMG) system for mouse click detection. The proposed VOG-EMG system combines gaze and muscle information to minimize the number of undesired clicks due to involuntary activations and environmental noise. We tested the system with 24 subjects, comparing three different configurations: one in which the VOG and EMG systems worked independently and two in which we used VOG gaze information to improve the EMG click detection. Results show that the number of false-positive click detections can be reduced when VOG and EMG information is combined. In addition, the third configuration, including extra processing, can reduce the activation delay produced because of the combined use of the VOG and EMG systems. The new VOG-EMG system is meant to be used in noisy environments in which the number of false clicks may impeach a reliable human-computer interaction.

Key words: eye tracking, gaze interaction, human-computer interface, Midas touch, motor control, multimodal interface, onset detection, rehabilitation, severe motor impairment, surface electromyography, video-oculography.

INTRODUCTION

Computers can improve the quality of life of people with severe motor-skill disabilities. Several applications have been developed to make their lives easier and more comfortable. Very often, alternative input devices other than the usual keyboard-mouse combination are required to interact with the computer. These include, among others, head-operated joysticks [1] and communication based on voice control [2].

However, cases exist in which all the aforementioned systems fail. For example, people in an advanced stage of amyotrophic lateral sclerosis have no control over certain muscle movements and therefore may have lost the ability to use a hand-controlled device such as a mouse [3]. Moreover, several motor impairments can arise because of other pathologies such as cerebral palsy, stroke, spinal lesions, and locked-in syndrome, among others [4]. In these cases, using eye movements such as a human-computer interaction technique is often the only solution [5]. The aim of a

gaze-tracking system is to use the eye gaze position (gaze point) of the user on the screen to control the user interface, i.e., the eye gaze substitutes for the computer mouse. For people with severe disabilities, this solution can be optimal. In fact, human-computer interaction for people with disabilities is one of the most popular applications of gaze-tracking systems [6]. The main techniques used for gaze tracking can be divided into two categories, depending on the technology used [7]:

1. Electro-oculography (EOG). It is based on placing a number of electrodes around the user's eye that measure voltage differences as the eye rotates inside its orbit. The signal detected depends on the movement of the eye and can be used to calculate gaze direction.

2. Video-oculography (VOG). It is based on image analysis, and it uses a camera to record images of the eye. The image changes as the eye moves, and the variation in the image features is used to deduce eye movement.

3. VOG has become a more popular technique than EOG mainly because of its lower intrusiveness. In addition, EOG systems are affected by interference from muscle signals and changes in pupil size.

VOG systems normally consist of one or more cameras and one or more infrared (IR) light sources. IR light provides a more stable illumination of the eye, allowing different eye features to be detected more robustly. Moreover, IR light produces reflections on the corneal surface that are visible in the image and can be used for calculating the gaze direction.

Among the existing VOG systems, two tendencies are found: high-resolution and low-resolution systems. The high-resolution system works with high-quality cameras and tries to accurately estimate gaze position on the screen [8–9]. On the other hand, low-resolution systems reduce the systems price by implementing gaze-tracking methods that use off-the-shelf components, such as Webcams [10–11]. These systems usually provide a lower accuracy than high-resolution systems; therefore, they need adapted interfaces, such as zooming or reducing the number of possible icons on the screen.

Gaze-tracking systems based on VOG were first developed in the '60s and '70s. However, the technology has significantly improved in the last decade. One of the first works published in eye tracking in 1974 already used a camera and IR lighting to track the user's gaze [12]. Today, eye gaze has been integrated in many commercial systems either as a communication device or as an eye-movement analysis tool, and most of these sys-

tems continue using IR light-emitting diodes (LEDs) and cameras in their hardware. Some issues still exist that prevent this technology from being useful for all potential users. Most gaze-tracking systems are sensitive to variations in light conditions, and their performance is greatly affected in outdoor scenarios. Furthermore, the user often needs to remain still in front of the camera to maintain the accuracy.

However, one of the most discussed topics in the eye-tracking community is the "Midas touch problem." This topic is defined as "the difficulty of determining intended activation of gazed features (the eyes do not register button clicks)" [7]. In other words, "everywhere you look, another command is activated; you cannot look anywhere without issuing a command" [13]. When the eye is moving around the screen, one cannot differentiate whether the user is just looking, is searching for something, or wants to select (click on) the gazed element. Many solutions have been proposed during the last few years to solve the Midas touch problem. The proposed alternatives can be classified in two groups:

1. Solutions based on eye actions: additional "eye actions" are required to infer an activation intention. These include blinking and maintaining the eye fixated on the desired element for a predefined "dwell time" [14–15].

2. Solutions based on multimodal interfaces: they combine gaze information with other supplementary user input. The idea is to use an alternative physiological signal from the user to infer the "click." Different possibilities are found in this group, such as brain computer interfaces, electromyography (EMG), and voice control [16–17].

The solutions of the first group have proven to be valid for several users who can maintain a stable fixation or can perform controlled blinks, even if it is not a "natural" task for the eye. However, users cannot always maintain a stable fixation. Moreover, although the gaze-based cursor movement has a lower pointing time (the cursor moves faster than the mouse because of the fast eye movements), the speed advantage is lost due to a higher selection time because dwell times usually range from 500 to 1,000 ms [13].

Multimodal interfaces are more versatile and can be adapted to each user according to his or her abilities. By gazing, a user moves the cursor, and by using an alternative input, such as sound (voice control), movement, or brain or muscle activity, a user makes the selection (i.e., the click).

255

NAVALLAS et al. Optimizing interoperability between VOG and EMG

EMG is the technique for evaluating and recording the electrical activity of muscles. Using muscle activity as an activation signal has proven to be feasible for many reasons. EMG systems can be adapted to the user in terms of the muscle selected and the sensitivity required to perform an activation. Any active muscle of the user can be used as input, and the system does not require body movement but rather the detection of a signal in the muscle, i.e., the system sensitivity can be adapted to the user's muscle activity, which results in a less-tiring system compared with other systems such as head buttons or sip/puff systems. In addition, EMG-based systems have been demonstrated to be faster and comparable with mouse selection speed [18].

However, EMG systems are affected by noise in the system. Depending on the signal strength, many involuntary selections can occur because of noise or involuntary muscle activations when high-sensitivity levels are used [19–20]. In addition, certain environments can also introduce artifacts due to movement. Recently, eye-tracking-based prototypes have been presented to drive wheelchairs [21–22]. The user uses gaze to drive the wheelchair with an interface in which the direction can be selected. The wheelchair will keep going in the same direction until a new selection is made. Undesired movements of the wheelchair can arise and are the main sources of involuntary selections; that is, they represent the "system noise."

This article presents a multimodal interface based on VOG and EMG for people who are severely disabled with eye-movement control and any kind of muscle activity. Different systems implementing these solutions can be found in the literature [18,23–25]. We present a novel VOG-EMG system with improved robustness to involuntary and noisy selections. The prototype introduced in this article aims to reduce the effect of the system noise with a communication protocol between the VOG and EMG systems. The system is described in the next section, which is divided into three subsections.

METHODS

VOG and EMG Systems

The system proposed in this article is a multimodal interface based on VOG and EMG systems. Essentially, the VOG system controls the movement of the mouse with the use of gaze, and the EMG system performs the selections with use of the frontalis muscle. The multimodal interface was developed to work in two different operation modes: (1) system with no communication (SNC) between the VOG and EMG systems (i.e., the usual operation mode) and (2) system with communication (SC) between the VOG and EMG systems, with gaze fixation detection for improving the EMG-driven selection detection. Two variants of the communication protocol have been developed.

The VOG and EMG systems are presented separately in the two sections that follow. The communication protocol between both systems is described in the third section.

Video-Oculographic System

The gaze-tracking system consists of a camera, working at 30 images/s at a resolution of 1024×768 pixels, and two groups of IR LEDs emitting at 880 nm (**Figure 1**). The camera is placed under the monitor, and the two LED groups are located at both sides of the screen. A focal length of 35 mm is used to capture a high-resolution image of the eye area. The camera mounts an optical filter that rejects light coming from external sources not in the IR band.

A gaze-tracking system consists of two main blocks: an eye-tracking block that extracts the information of the eye features from the image and a gaze-estimation block that uses these eye features to estimate where the user is looking on the screen. In the system described in this article, the eye-tracking (or image-processing) block uses the eye image as input and processes it to calculate the position of the pupil center and the two glints. Image segmentation is done with the use of information on gray level, size, compactness, and glint-pupil relative positions. In the first image, a search is conducted in the whole image. For subsequent images, the information of the previous image is used for defining a search window. If the pupil and glints are not found in the window, the system searches the whole image again. In the gaze-estimation block, the glints and pupil positions are used as input to the function that estimates the gaze-point coordinates in the screen. Generally, a second- or third-order polynomial is selected as the gaze estimation function based on unknown coefficients [26–27] in which glints and pupil center positions are used as input variables:

$$(p_x, p_y) = g(\bar{c}, \bar{f}),$$

where p_x and p_y are the coordinates of the gaze point on the screen, $g(\cdot)$ is the gaze estimation function, $\bar{c}$ is the vector of unknown coefficients, and $\bar{f}$ is the vector of features extracted from the image.
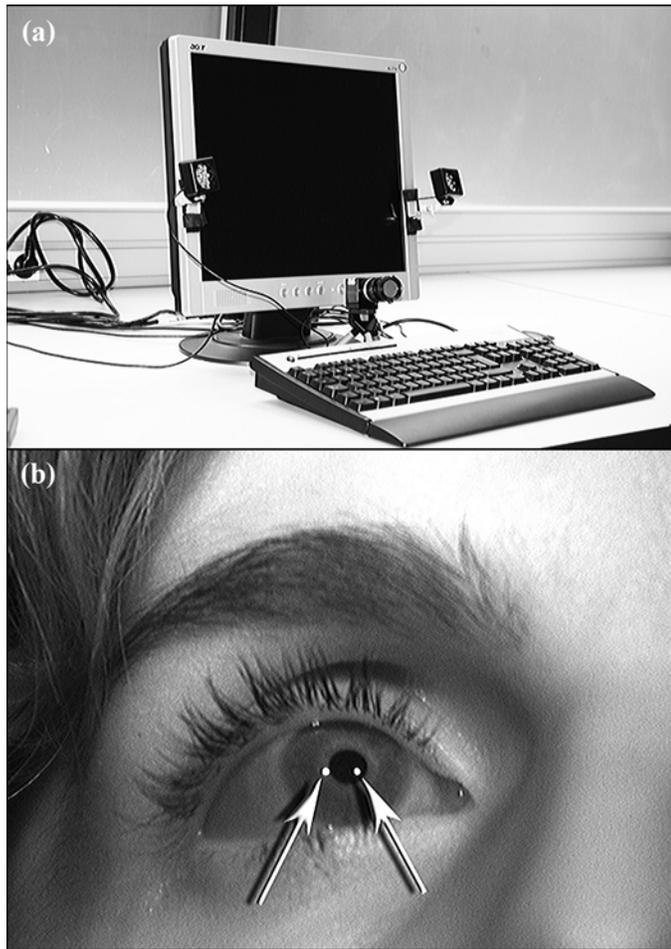
Figure 1.
(a) Video-oculography system composed of monitor, camera, and two groups of infrared light-emitting diodes located at both sides of screen. **(b)** Image recorded by camera. Two bright dots, indicated by arrows, are reflections, i.e., glints, produced by light sources on corneal surface.

The coefficients of this function are determined by means of a calibration procedure. The user is asked to gaze at certain points on the screen. Normally, a 3×3 or 4×4 grid of points uniformly distributed on the screen is used. The user gazes at each of the points while the system records the pupil and glints positions from the images associated with the calibration point coordinates. Once the grid is completed, the system calculates the unknown coefficients of the selected gaze-estimation function using the calibration information. This function is then interpolated for the rest of the points on the screen with similar accuracy. The use of the operating system (OS) application programming interface moves the mouse. The gaze-tracking system calls this function each time a screen point is calculated to perform cursor movement.

The gaze-tracking system records eye movement to differentiate between fixations and saccadic movements. A fixation of the eye is defined as a quasistable position of the eye in an area <1° of the visual angle [28]. Saccadic movements are high-speed jumps between fixations. The velocity of a saccade depends linearly on the amplitude, e.g., a 10° amplitude is associated with a velocity of 300°/s, and 30° is associated with 500°/s. On the other hand, we can characterize a fixation by applying explicit spatial and temporal criteria. We can detect a fixation if the eye remains within a radius of 1° during a temporal fixation threshold of 200 ms, which is the de facto standard [19]. The gaze-tracking system measures eye-movement spatial dispersion and speed to infer fixation status (**Figure 2(a)**) and differentiate between fixation and saccadic movements of the eye.

Electromyographic System

EMG deals with the acquisition and processing of the electrical activity associated with muscle contraction. EMG techniques can be divided into two groups: intramuscular EMG and surface EMG. Intramuscular EMG, which
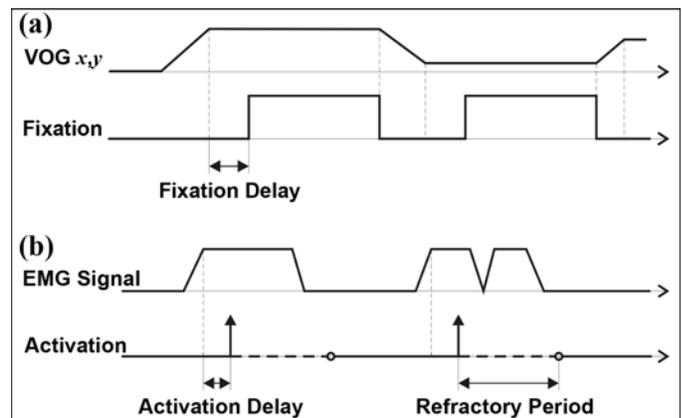

Figure 2.
Schematics that represent behavior of video-oculography (VOG) and electromyographic (EMG) systems: **(a)** VOG system outputs are (x, y) coordinates of gaze within screen (shown only in one dimension for sake of clarity) and Boolean flag indicating fixation. Fixation delay is minimum time starting at physiological fixation required by system to activate fixation flag. **(b)** EMG signal is processed by EMG system for determining activation of muscle. Activation delay is processing delay of system, and refractory period is temporal window in which EMG system may not generate any additional activation signal.

257

NAVALLAS et al. Optimizing interoperability between VOG and EMG

requires the insertion of an indwelling electrode (usually a needle electrode), is employed in the assessment of neuromuscular diseases in clinical practice [29]. Surface EMG uses one or more electrodes attached to the skin covering the muscle under study and has applications in nerve conduction studies [30], ergonomics [31], kinesiology [32], prostheses control [33], and biofeedback [34]. Hence, we have chosen surface EMG recording to assess the activation of a given muscle because of its low invasiveness and good performance. A typical EMG system consists of a set of recording electrodes, a signal amplifier, a data acquisition unit, and postprocessing software. In our EMG system, two disposable, self-adhesive pre-gelled electrodes are attached to the forehead, over the frontalis muscle, and a third one is attached to the wrist. The exact position of the electrodes is not critical to the setup. The signals from the two electrodes on the forehead are the input to a differential amplifier (bipolar EMG recording configuration [35]), while the signal from the electrode on the wrist acts as a reference. The preprocessing stage of the signal amplifier includes a band-rejection filter, tuned to 50 Hz, and a bandpass filter with cutoff frequencies of 5 and 500 Hz. The amplified signal is digitized (32 b/sample, 2,400 samples/s) and transmitted to a personal computer. The software for the postprocessing and user interface tools was programmed in LabVIEW (National Instruments Corporation; Austin, Texas). This program includes the algorithm for the muscle activation detection that determines whether the muscle is activated or not.

Some techniques analyze the signal in the frequency domain to detect muscle activation, while others limit their analysis to the temporal domain. We have focused our attention on temporal analysis because the computational complexity is reduced. Most detection algorithms consist of up to three basic processing stages, namely signal conditioning, detection, and postprocessing. In general, all systems establish a threshold value that should be exceeded by a function of the input samples for determining that an activation has occurred. We selected the activation detection algorithm after an evaluation study of some of the available algorithms. We compared six different algorithms, namely, Hodges and Bui [36], Bonato et al. [19], Lidierth [37], Abbink et al. [38], and approximated generalized likelihood ratio (AGLR) step and AGLR ramp [39].

The evaluation study for selecting the activation detection algorithm is detailed elsewhere [40]. In brief, we tested the algorithms in 50 series of experiments (10 users/5 types of experiments) in which each user had to activate the muscle after a visual stimulus. We presented 100 stimuli in each experiment. The different series included different muscles and different levels of voluntary muscle contraction. Each series provided one EMG signal that we used to evaluate the six algorithms offline, given that the expected activation pattern was known in advance. We measured the quality of the algorithms with a figure of merit defined as a function of the number of false-positive detections and the number of false-negative detections and measured with the processing delay introduced by the algorithm. The results showed that the AGLR step is the best balance between the figure of merit and the processing delay.

The AGLR-step algorithm calculates an estimate of the muscle activity as a function of the mean and variance of the level of activity at rest and a set of EMG samples within a 9.60 ms window [39]. Whenever this activity estimate exceeds a predefined threshold value, the activation flag is set to "true." This threshold value is determined during the calibration of the algorithm. In the final implementation of the activation detection algorithm, the signal is analyzed within the time between buffers (8.33 ms), leading to a processing delay of between 8.33 and 16.67 ms. An activation flag is set to "true" whenever the AGLR-step algorithm detects muscle activity. If the EMG system is operating without communication with the VOG system, a mouse click event is sent to the OS when the activation flag is set to "true." Additionally, a refractory period of 200 ms is established, during which the signal is not analyzed by the AGLR-step algorithm (an ongoing activation associated with the click is assumed to occur during the refractory period) and after which the activation flag returns to "false," and the signal analysis restarts (**Figure 2(b)**).

Multimodal Interface VOG+EMG

Once we implemented and tested both systems, we combined them in a single interface (VOG+EMG). The gaze-tracking system was used to move the cursor on the screen, while the EMG system used muscle activation as a click signal and sent a mouse click event to the OS. We implemented three different multimodal systems:

1. SNC: VOG and EMG work independently. Each time a muscle activation is detected, a click event is sent to the OS, regardless of the VOG status (**Figure 3(a)**). A multimodal system usually works this way.

2. SC: A communication protocol is established between both systems. The EMG system receives information
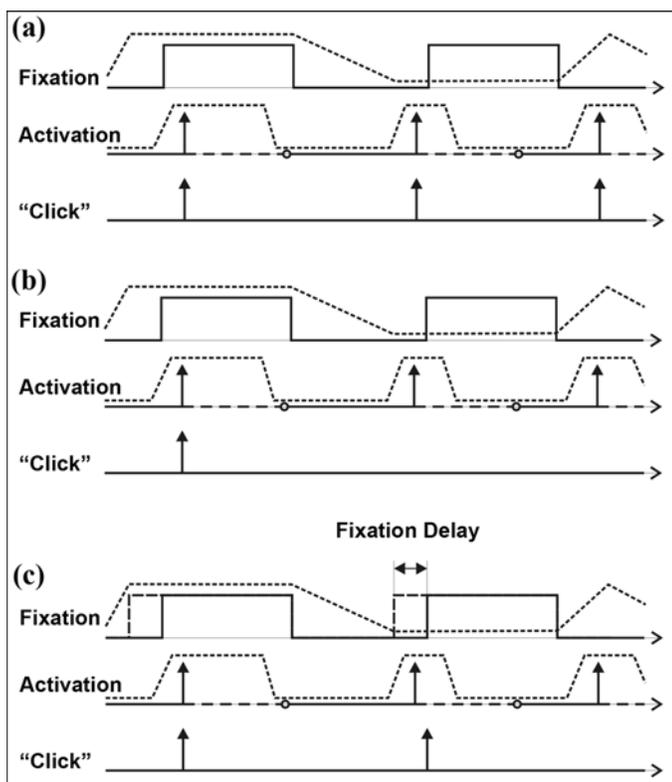
Figure 3.
Schematics that represent behavior of three different system approaches to video-oculography (VOG)+ electromyographic (EMG) multimodal interaction: Systems with **(a)** no communication in which all activations detected by EMG system generate click event, **(b)** communication in which only activations detected when fixation is present generate click event, and **(c)** communication and fixation delay correction in which activations occurring within temporal window before fixation are detected by VOG system also generate click event that is sent to operating system as soon as VOG reports fixation.

about the fixation status of the eye from the VOG system. When an activation is detected, the EMG system sends a click event to the system only if a simultaneous fixation occurs (**Figure 3(b)**).

3. System with communication and fixation delay correction (SCDC): The EMG system also receives information about the fixation status of the eye from the VOG system. When an activation is detected, the EMG system issues a click event only if a simultaneous fixation occurs. Additionally, when the EMG receives the fixation signal, it checks whether an activation has occurred in a temporal window before that moment. The length of the temporal window is equal to the fixa-

tion delay, and the click is not sent until the fixation is achieved (**Figure 3(c)**).

Our hypothesis is that voluntary activations only occur in maintained fixation situations. In other words, muscle activations that occur when the eye is not fixating (e.g., performing saccadic movements) can be considered involuntary and therefore rejected. In the second system (SC), once an activation is detected, the EMG system will check the eye fixation status and will only issue a click event to the OS if a fixation is occurring. This procedure might introduce a delay on the click signaling in situations in which the muscle activation is performed before the VOG system has reported a fixation. In the third system (SCDC), an attempt is made to compensate for this delay in the fixation detection. Activations are accepted that occur between the moment the fixation starts and the moment it is detected by the gaze-tracking system. However, the click event is not issued until the VOG reports a fixation.

Activations in no-fixation situations can happen because of involuntary user actions (e.g., spasmodic activation) or system noise. In both cases, these undesirable activations can be modeled by introducing noise to the system. The experiments performed to compare the three systems (SNC, SC, and SCDC) are described in the next section.

Experimental Protocol

Participants and Apparatus

Twenty-four volunteers, ranging from 24 to 41 years old, participated in the study. All subjects were nondisabled with no diagnosed neurological impairment and signed the informed consent. They were randomly divided into three groups of eight people each: one group tested the SNC, another group tested SC, and the other group tested SCDC. Three users had previous experience with gaze tracking and the EMG selection method, and each of them was assigned to a different group. We used a 17 in. monitor with a resolution of $1,280 \times 1,024$ to present the target-acquisition task. The gaze tracker and the EMG system introduced previously were used in the experiment. **Figure 4** shows the experimental setup with one of the participants during the test.

Design and Procedure

We conducted the experiment using a $3 \times 3 \times 2$ mixed design. Factors were type of communication (SNC, SC, or

259

NAVALLAS et al. Optimizing interoperability between VOG and EMG

Figure 4.
Experimental setup. Main screen is displaying target-acquisition task, while secondary monitor is displaying fixation state and electromyographic signal.

SCDC) between subjects, noise level (no noise [NN], low noise [LN], or high noise [HN]), and fixation delay (200 ms or 400 ms) within subjects.

The experiment consisted of a target-acquisition task as specified in the International Organization for Standardization standard 9241-9, following a similar procedure as in the study by Zhang and MacKenzie [41]. Sixteen targets were arranged in a circular layout, and participants had to point at each target using their eyes and select it by a frowning or tightening their jaws. The participants were instructed to select the targets as accurately as possible. The size of the targets was fixed to 150 pixels in diameter. Prior to starting the experiment, participants calibrated both systems and ran a warm-up trial to become acquainted with the multimodal interface. The warm-up trial consisted of 64 targets for each subject, with a fixation delay of 200 ms and no artificial noise added. Each participant completed a block of 64 trials (four times each target in a random sequence) for each combination of noise level and fixation delay, using the type of communication between the gaze tracker and EMG corresponding to his or her group. The order of the six blocks was counterbalanced to neutralize learning effects.

In a normal situation, a user usually performs two types of tasks: a visual task, such as browsing or reading, and a selection task to activate a menu item or a link. During the visual task, no clicks should be performed because they can lead to an undesired selection. We added a moving object to simulate the situation in which the user is not fixating and does not want to activate a click. The object

appeared randomly between 2 and 4 s and followed a lemniscate of Bernoulli curve with constant speed. Once the random time elapsed, this object disappeared and a new target that the user had to select was displayed. Participants were instructed not to perform any activation while the moving object was on the screen.

In each trial, we measured completion time (i.e., time required to select the target since the moment it appeared), error rate (i.e., the proportion of targets selected when the cursor was outside the target), involuntary selections (i.e., the proportion of targets selected by noise and not by an EMG activation), and clicks produced by noise (i.e., the proportion of noisy activations that issued a click event).

Simulation of Noisy Environment

Noisy environments in which involuntary muscle activations occur might lead to undesired selections when employing an EMG system to perform clicks. For example, the movement of a wheelchair will introduce noise in the system, thus generating undesired activations. People with involuntary facial movements will also perform undesired selections due to involuntary activation of the targeted muscle or to cross talk from involuntary activation of nearby muscles.

We stimulated a noisy environment in which undesired muscle activations occur by introducing random activations to the EMG system. A homogeneous Poisson process was chosen to model the noisy activations. In the LN condition, the average time between noisy activations (the inverse of the intensity of the Poisson process) was 8 s, while in the HN condition, it was 4 s. The noisy activations were treated the same as the voluntary ones regarding the multimodal system.

RESULTS

Results are presented separately for all four dependent variables of the experiment: completion time, error rate, number of involuntary selections, and number of noisy clicks for each target. For each measured variable, the analysis is divided into two blocks: the results obtained with the fixation delay set to 200 ms and the results obtained with the fixation delay set to 400 ms. Fixation delay did not influence the results of SNC; therefore, only one fixation delay was used in that condition.

Completion Time

Completion time was defined as the mean time used by the user since the target was displayed until the selection was performed. We calculated this mean time by averaging all the targets of each trial. **Figure 5** shows the results for each type of communication and noise level for both fixation delays.

The type of communication used affected completion time significantly, $F_{2,21} = 41.01$, $p < 0.05$. The use of communication (SC and SCDC) increased the completion time, with the increase higher in the 400 ms fixation delay than in the 200 ms. This delay was partially compensated when the fixation delay correction was incorporated (SCDC). SC and SCDC were compared, and SCDC resulted in a faster system ($p < 0.05$ after Bonferroni adjustment). Noise did not significantly affect the completion time of any of the systems, $F_{2,42} = 1.72$, $p > 0.05$.
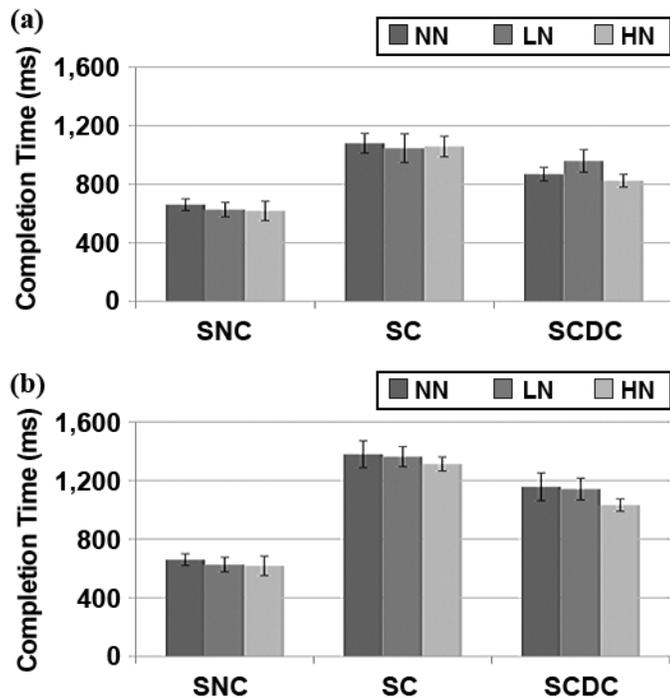
Error Rate

The error rate was defined as the ratio between the number of targets selected when the cursor was outside the target (unsuccessful activations) and the total number of targets for each trial. **Figure 6** shows the error rate measured for each type of communication, noise level, and fixation delay.

The type of communication used significantly affected error rate, $F_{2,21} = 19.66$, $p < 0.05$. The SC and SCDC presented a lower error rate than that for the SNC. Noise level also significantly affected error rate, $F_{2,42} = 6.53$, $p < 0.05$. A post hoc analysis with Bonferroni adjustment showed that error rate was significantly higher for the HN condition. Fixation delay significantly affected error rate, $F_{1,21} = 4.94$, $p < 0.05$. Error rates for SC and SCDC were slightly reduced when a fixation delay of 400 ms was used instead of 200 ms.
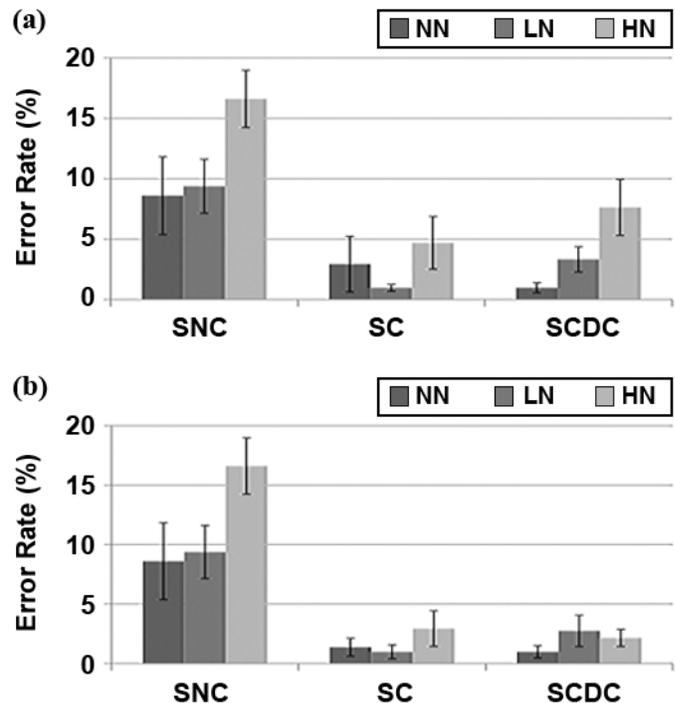

Figure 5.
Completion time obtained with fixation delay of **(a)** 200 and **(b)** 400 ms. In each plot, three systems (with no communication [SNC], with communication [SC], and with communication and fixation delay correction [SCDC]) are tested in three noise conditions (no noise [NN], low noise [LN], and high noise [HN]). Mean value (bars) and standard error of the mean (error bars) are depicted.

Figure 6.
Error rate obtained with fixation delay of **(a)** 200 and **(b)** 400 ms. In each plot, three systems (with no communication [SNC], with communication [SC], and with communication and fixation delay correction [SCDC]) were tested in three noise conditions (no noise [NN], low noise [LN], and high noise [HN]). Mean value (bars) and standard error of mean (error bars) are depicted.

261

NAVALLAS et al. Optimizing interoperability between VOG and EMG

Involuntary Selections

The number of involuntary selections was defined as the number of targets selected (click over target) due to noise and not due to EMG activation. For all the experiments, the number of involuntary selections increased with noise, $F_{2,42} = 107.51$, $p < 0.05$. The real noise level was low (signal-to-noise ratio 37 to 42 dB), ensuring that the number of involuntary activations due to real noise was negligible. This increase was especially noticeable in SCDC with a 200 ms fixation delay. SCs had a significantly lower number of involuntary selections, $F_{2,21} = 13.61$, $p < 0.05$, except for SCDC with a 200 ms fixation delay. Apart from this finding, the remaining systems were not significantly different differences between the 200 ms and 400 ms activation windows.

Noisy Clicks for each Target

The number of noisy clicks for each target was defined as the number of clicks (not necessarily selections) due to noise that occurred during the presentation of a single target. This result included the time during which the moving target was displayed, i.e., before the current target was presented. **Figure 7** depicts the noisy clicks for each target measured for each type of communication, noise level, and fixation delay.

The type of communication significantly affected the ratio of noisy clicks for each target, $F_{2,21} = 369.09$, $p < 0.05$, and SC and SCDC performed considerably better than SNC, with a reduction of the number of noisy clicks by more than four-fold. No significant differences between SC and SCDC were present with a fixation delay of 200 ms and with 400 ms. Fixation delay significantly affected noisy clicks for each target, $F_{1,21} = 23.26$, $p < 0.05$. In **Figure 7,** a reduction of the noisy clicks for each target is shown for a fixation delay of 400 ms as compared with 200 ms for both systems with communication. Noise level also has a significant effect, $F_{2,42} = 387.72$, $p < 0.05$, with noisy clicks increasing with noise.

DISCUSSION

The use of information on eye fixations in detecting voluntary facial-muscle selections improved the robustness of multimodal interfaces to noise based on VOG and EMG, at the expense of a slight increase of selection time. Specifically, the VOG-EMG combination proposed
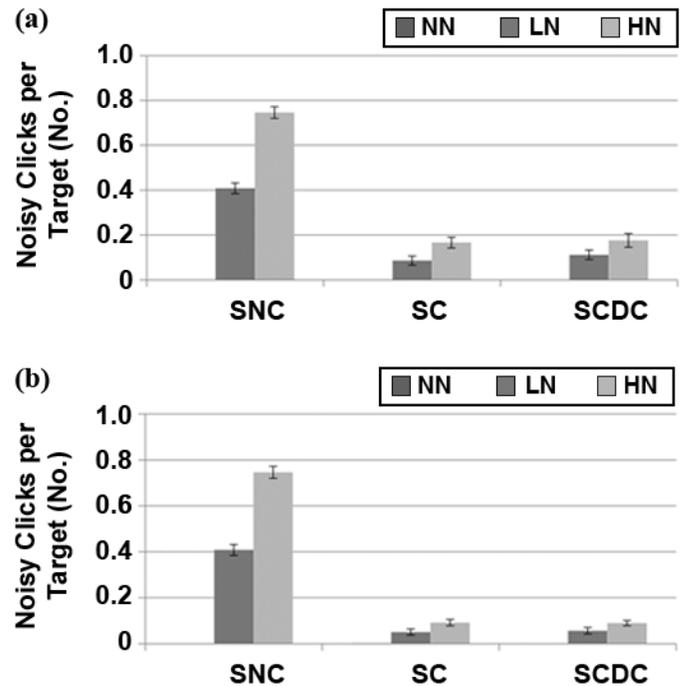

Figure 7.
Noisy clicks for each target obtained with fixation delay of **(a)** 200 and **(b)** 400 ms. In each plot, three systems (with no communication [SNC], with communication [SC], with communication and fixation delay correction [SCDC]) were tested in three noise conditions (no noise [NN], low noise [LN], and high noise [HN]). Mean value (bars) and standard error of mean (error bars) are depicted.

in this article helped reduce the error rate due to system noise by requiring a gaze fixation to issue a click event when a muscle activation was detected. Therefore, the number of potential erroneous selections during a visual task in which the user did not intend to perform any activation was reduced. As a result, the robustness of the system increased compared with that of a system in which this tight integration between the VOG and EMG systems did not exist.

The obtained error rate for the SNC in NN conditions (12%) was slightly below the error rates obtained in previous studies: 17 [42], 24 [18], and 34 percent [43]. When communication between the VOG and EMG systems was used, error rates fell below 5 percent in NN and LN environments and below 10 percent in the HN environment. These low error rates, even without communication, may be because we fully controlled the operation of both subsystems separately. Specifically, the VOG system included a fast fixation estimation algorithm (described in the section "Video-Oculographic System," page 255) that made the cursor follow gaze direction more closely com-

pared with other available gaze-tracking systems. On the other hand, the EMG system employed a statistical estimation method for activating detection [39], which performed more accurately than simple thresholding algorithms.

Communication was delayed because the fixation condition could not be detected instantly. In fact, a fixation is defined as a quasistable position of the eye (spatial dispersion below 1°) during at least 150 ms [28]. In addition, for the activation flag to be raised, the system required a sustained fixation during the predefined fixation delay (200 and 400 ms). These findings imply a minimum unavoidable delay for the proposed system, as shown in the results for the completion time when SNC and SC frameworks were compared. The SCDC tried to compensate for this delay by accepting muscle activations that occurred during the fixation delay (while the fixation was occurring but still not detected). As the results for completion time for this system show (**Figure 5**), part of the additional delay introduced by SC was compensated. However, the mean completion time was still higher than the completion time of SNC.

SCDC, however, presented an undesired effect especially noticeable in HN conditions when involuntary selections are measured with 200 ms as the fixation delay. As shown in **Figure 8**, the number of involuntary selections equaled the rate of the SNC system. This effect was not present when the fixation delay was set to 400 ms, where the number of involuntary selections was almost equal to that of SC. This effect may be produced by the way the number of involuntary selections was defined: whenever a selection was made, the evaluation protocol looked for a voluntary activation within the fixation delay before the fixation, disregarding any involuntary selection due to noise. Only if a voluntary selection was not found was the event declared involuntary. Thus, when the activation was set to 400 ms, the probability of finding a voluntary activation was doubled compared with that of 200 ms.

In summary, SCDC performed adequately in NN and LN conditions, with involuntary selections and noise errors compared with those of SC. However, it presented lower completion time than SC, which resulted in a faster interface. On the other hand, SC was more accurate in HN environments, at the expense of a slight increase in processing delay, as demonstrated by the completion time results. Hence, we suggest the possibility of having the fixation delay compensation as an additional mechanism that may be switched off whenever the external conditions may induce HN in the system.
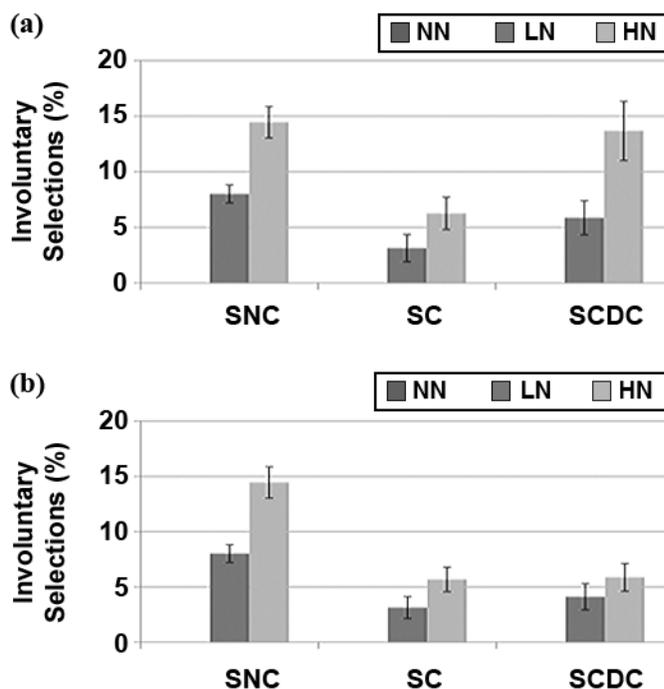

Figure 8.

Involuntary selections obtained with fixation delay of **(a)** 200 and **(b)** 400 ms. In each plot, three systems (with no communication [SNC], with communication [SC], and with communication and fixation delay correction [SCDC]) were tested in three noise conditions (no noise [NN], low noise [LN], and high noise [HN]). Mean value (bars) and standard error of mean (error bars) are depicted.

Further experiments are needed to validate the system with subjects with motor difficulties or neurological impairments for which the casuistry is wider and the tests are difficult to standardize. Our first objective was to validate the system and our hypothesis with a high number of nondisabled individuals so that we can obtain statistically significant results. Experts working with users who are disabled should be recruited to plan the experiments and the tests and to develop practical applications of the interface in the fields of communications and rehabilitation. We consider that highly valuable information can be obtained from the studies with users who are disabled in which rate versus accuracy trade-off can be considered. While many of the EMG-based systems try to increase communication speed, we expect that increasing the accuracy can decrease the fatigue and frustration for many users who are disabled.

263

NAVALLAS et al. Optimizing interoperability between VOG and EMG

CONCLUSIONS

A multimodal interface that combines gaze pointing and EMG clicking has been implemented and evaluated. Both subsystems were designed independently and combined by means of a communication protocol, allowing for a higher robustness to noise: the use of eye fixation information provided by the gaze tracker in detecting muscle activations improved the performance of the whole interface in terms of undesired activations. However, completion time was increased because of a simultaneous fixation to issue a click event.

The tight integration between the gaze tracker and EMG system proposed in this article reduced the number of undesired selections that occurred when the user performed a visual task such as reading or browsing. This reduction might benefit environments in which false activations likely occur, for example, when the user is driving a wheelchair or has involuntary facial muscle movements, hence increasing the robustness and reliability of the multimodal interface.

ACKNOWLEDGMENTS

REFERENCES

1. Evans DG, Drew R, Blenkhorn P. Controlling mouse pointer position using an infrared head-operated joystick. IEEE Trans Rehabil Eng. 2000;8(1):107–17. [PMID: 10779114] DOI:10.1109/86.830955

2. Su MC, Chung MT. Voice-controlled human-computer interface for the disabled. Comput Control Eng J. 2001; 12(5):225–30. DOI:10.1049/cce:20010504

3. Calvo A, Chiò A, Castellina E, Corno F, Farinetti L, Ghiglione P, Pasian V, Vignola A. Eye tracking impact on quality-of-life of ALS patients. Proceedings of the 11th International Conference on Computers Helping People with Special Needs; 2008 Jul 9–11; Linz, Austria. Berlin (Germany): Springer-Verlag. 2008. p. 70–77.

4. Donegan M. Eye control hints and tips. Communication by gaze interaction (COGAIN), IST-2003-511598: Deliverable 3.4. Online multimedia training resource to help in the initial assessment process and eye control take-up; 2007.

5. Donegan M. D3.5 Exemplar eye control activities for users with complex needs (COGAIN), IST-2003-511598: Deliverable 3.5. Freely available exemplar eye control activities for end-users with complex needs; 2008.

6. Majaranta P, Räihä KJ. Twenty years of eye typing: Systems and design issues. Proceedings of the 2002 Symposium on Eye Tracking Research & Applications; 2002 Mar 25–27; New Orleans, LA. New York (NY): ACM; 2002. p. 15–22.

7. Duchowski AT. Eye tracking methodology: Theory and practice. 2nd ed. New York (NY): Springer; 2007.

8. Model D, Eizenman M. User-calibration-free remote gaze estimation system. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications; 2010 Mar 22–24; Austin, TX. New York (NY): ACM; 2010. p. 29–36.

9. Villanueva A, Cabeza R. A novel gaze estimation system with one calibration point. IEEE Trans Syst Man Cybern B Cybern. 2008;38(4):1123–38. [PMID: 18632402] DOI:10.1109/TSMCB.2008.926606

10. San Agustín J, Skovsgaard H, Hansen JP, Hansen DW. Low-cost gaze interaction: Ready to deliver the promises. Proceedings of the 27th International Conference on Human Factors in Computing Systems; 2009 Apr 4–9; Boston, MA. New York (NY): ACM; 2009. p. 4453–58.

11. Li D, Babcock J, Parkhurst DJ. openEyes: A low-cost head-mounted eye-tracking solution. Proceedings of the 2006 Symposium on Eye-Tracking Research & Applications; 2006 Mar 27–29; San Diego, CA. New York (NY): ACM; 2006. p. 95–100.

12. Merchant J, Morrissette R, Porterfield JL. Remote measurement of eye direction allowing subject motion over one cubic foot of space. IEEE Trans Biomed Eng. 1974;21(4): 309–17. [PMID: 4837476] DOI:10.1109/TBME.1974.324318

13. Jacob RJ. What you look at is what you get: Eye movement-based interaction techniques. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People; 1990 Apr 1–5; Seattle, WA. New York (NY): ACM; 1990. p. 11–18.

14. MacKenzie S, Ashtiani B. BlinkWrite: Efficient text entry using eye blinks. Univ Access Inf Soc. 2010:5–13.

15. Majaranta P, Ahola UK, Špakov O. Fast gaze typing with an adjustable dwell time. Proceedings of the 27th International Conference on Human Factors in Computing Systems; 2009 Apr 4–9; Boston, MA. New York (NY); 2009. p. 357–60.

16. Karray F, Alemzadeh M, Saleh JA, Arab MN. Human-computer interaction: Overview on state of the art. Int J Smart Sens Intell Syst. 2008;1(1):137–59.

17. Porta M. Human-computer input and output techniques: An analysis of current research and promising applications. Artif Intell Rev. 2007;28(3):197–226.
[DOI:10.1007/s10462-009-9098-5](DOI:10.1007/s10462-009-9098-5)

18. Mateo JC, San Agustín J, Hansen JP. Gaze beats mouse: Hands-free selection by combining gaze and EMG. Proceedings of the CHI '08 Extended Abstracts on Human Factors in Computing Systems; 2008 Apr 5–10; Florence, Italy. New York (NY): ACM; 2008. p. 3039–44.

19. Bonato P, D'Alessio T, Knaflitz M. A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. IEEE Trans Biomed Eng. 1998;45(3):287–99. [PMID: 9509745]
[DOI:10.1109/10.661154](DOI:10.1109/10.661154)

20. Merlo A, Farina D, Merletti R. A fast and reliable technique for muscle activity detection from surface EMG signals. IEEE Trans Biomed Eng. 2003;50(3):316–23.
[PMID: 12669988]
DOI:10.1109/TBME.2003.808829

21. Gajwani PS, Chhabria SA. Eye motion tracking for wheelchair control. Int J Inform Technol Knowl Manag. 2010; 2(2):185–87.

22. Purwanto P, Mardiyanto R, Arai K. Electric wheelchair control with gaze direction and eye blinking. Artif Life Robot. 2009;14(3):397–400. [DOI:10.1007/s10015-009-0694-x](DOI:10.1007/s10015-009-0694-x)

23. Chin CA, Barreto A, Cremades JG, Adjouadi M. Integrated electromyogram and eye-gaze tracking cursor control system for computer users with motor disabilities. J Rehabil Res Dev. 2008;45(1):161–74. [PMID: 18566935]
DOI:10.1682/JRRD.2007.03.0050

24. Junker AM, Hansen JP. Gaze pointing and facial EMG clicking. Proceedings of the 2nd COGAIN Conference; 2006 Sep 4–5; Turin, Italy. Copenhagen (Denmark): COGAIN. p. 40–43.

25. Chin CA, Barreto A. Enhanced hybrid electromyogram/eye gaze tracking cursor control system for hands-free computer interaction. Proceedings of the 28th IEEE EMBS Annual International Conference; 2006 Aug 30–Sep 3; New York, NY. Piscataway (NJ): IEEE EMBS. p. 2296–99.

26. Morimoto CH, Mimica MR. Eye gaze tracking techniques for interactive applications. Comput Vis Image Underst. 2005;98(1):4–24. DOI:10.1016/j.cviu.2004.07.010

27. Cerrolaza JJ, Villanueva A, Cabeza R. Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. Proceedings the 2008 Symposium on Eye-Tracking Research & Applications; 2008 Mar 26–28; Savannah, GA. New York (NY): ACM; 2008. p. 259–66.

28. Salvucci DD, Goldberg JH. Identifying fixations and saccades in eye-tracking protocols. Proceedings of the 2000 Symposium on Eye-Tracking Research & Applications; 2000 Nov 6–9; Palm Beach Gardens, FL. New York (NY): ACM; 2000. p. 71–78. [DOI:10.1145/355017.355028](DOI:10.1145/355017.355028)

29. Stålberg E, Daube J. Electromyographic methods. In: Stålberg E, editor. Clinical neurophysiology of disorders of muscle and neuromuscular junction, including fatigue. Amsterdam (the Netherlands): Elsevier; 2003. p. 147–86.

30. Falck B. Neurography—Motor and sensory nerve conduction studies. In: Stålberg E, editor. Clinical neurophysiology of disorders of muscle and neuromuscular junction, including fatigue. Amsterdam (the Netherlands): Elsevier; 2003. p. 269–322.

31. Hagg GM, Melin B, Kadefors R. Application in ergonomics. In: Merletti R, Parker P, editors. Electromyography: Physiology, engineering, and noninvasive applications. Hoboken (NJ): John Wiley & Sons, Inc; 2004. p. 343–64.
[DOI:10.1002/0471678384.ch13](DOI:10.1002/0471678384.ch13)

32. Frigo C, Shiavi R. Applications in movement and gait analysis. In: Merletti R, Parker P, editors. Electromyography: Physiology, engineering, and noninvasive applications. Hoboken (NJ): John Wiley & Sons, Inc; 2004. p. 381–402.
[DOI:10.1002/0471678384.ch15](DOI:10.1002/0471678384.ch15)

33. Parker P, Englehart KB, Hudgins BS. Control of powered upper limb prostheses. In: Merletti R, Parker P, editors. Electromyography: Physiology, engineering, and noninvasive applications. Hoboken (NJ): John Wiley & Sons, Inc; 2004. p. 456–76. [DOI:10.1002/0471678384.ch18](DOI:10.1002/0471678384.ch18)

34. Cran JR. Biofeedback applications. In: Merletti R, Parker P, editors. Electromyography: Physiology, engineering, and noninvasive applications. Hoboken (NJ): John Wiley & Sons, Inc; 2004. p. 435–52.[DOI:10.1002/0471678384.ch17](DOI:10.1002/0471678384.ch17)

35. Merletti R, Hermens HJ. Detection and conditioning of the surface EMG signal. In: Merletti R, Parker P, editors. Electromyography: Physiology, engineering, and noninvasive applications. Hoboken (NJ): John Wiley & Sons, Inc; 2004. p. 107–32.[DOI:10.1002/0471678384.ch5](DOI:10.1002/0471678384.ch5)

36. Hodges PW, Bui BH. A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. Electroencephalogr Clin Neurophysiol. 1996;101(6):511–19. [PMID: 9020824]

37. Lidierth M. A computer based method for automated measurement of the periods of muscular activity from an EMG and its application to locomotor EMGs. Electroencephalogr Clin Neurophysiol. 1986;64(4):378–80. [PMID: 2428587]
[DOI:10.1016/0013-4694(86)90163-X](DOI:10.1016/0013-4694(86)90163-X)

38. Abbink JH, Van der Bilt A, Van der Glas HW. Detection of onset and termination of muscle activity in surface electromyograms. J Oral Rehabil. 1998;25(5):365–69.
[PMID: 9639161]
[DOI:10.1046/j.1365-2842.1998.00242.x](DOI:10.1046/j.1365-2842.1998.00242.x)

39. Staude G, Flachenecker C, Daumer M, Wolf W. Onset detection in surface electromyographic signals: A systematic

265

NAVALLAS et al. Optimizing interoperability between VOG and EMG

comparison of methods. EURASIP J Appl Signal Process. 2001;(11):67–81. [DOI:10.1155/S1110865701000191](#)

40. Ariz M, Navallas J, Villanueva A, San Agustín J, Cabeza R, Tall M. D4.9 Online information on how to use control facilities to supplement gaze for control. 2009. Communication by Gaze Interaction (COGAIN). Available from: http://www.cogain.org/wiki/COGAIN_Reports.

41. Zhang X, MacKenzie IS. Evaluating eye tracking with ISO 9241-Part 9. Proceedings of the 12th International Conference on Human-Computer Interaction: Intelligent Multimodal Interaction Environments; 2007 Jul 22–27; Beijing, China. New York (NY): ACM; 2007.

42. Surakka V, Illi M, Isokoski P. Gazing and frowning as a new human-computer interaction technique. ACM Trans Appl Percept. 2004;1(1):40–56. [DOI:10.1145/1008722.1008726](#)

43. Partala T, Aula A, Surakka V. Combined voluntary gaze direction and facial muscle activity as a new pointing technique. Proceedings of INTERACT; 2001 Jul 9–13; Tokyo, Japan. Amsterdam (the Netherlands): IOS Press; 2001. p. 100–107.