Applied Soft Computing 20 (2014) 95–102

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Classification of silent speech using support vector machine and relevance vector machine

Mariko Matsumoto[a,*], Junichi Hori[b]

[a] Graduate School of Science & Technology, Niigata University, Niigata, Japan
[b] Department of Biocybernetics, Niigata University, Niigata, Japan

ARTICLE INFO

Article history:
Received 16 March 2013
Received in revised form 19 July 2013
Accepted 10 October 2013
Available online 21 November 2013

Keywords:
EEG
Silent speech
Brain computer interface (BCI)
Support vector machine (SVM)
Relevance vector machine (RVM)
Adaptive collection

ABSTRACT

To provide speech prostheses for individuals with severe communication impairments, we investigated a classification method for brain computer interfaces (BCIs) using silent speech. Event-related potentials (ERPs) were recorded using scalp electrodes when five subjects imagined the vocalization of Japanese vowels, /a/, /i/, /u/, /e/, and /o/ in order and in random order, while the subjects remained silent and immobilized.

For actualization, we tried to apply relevance vector machine (RVM) and RVM with Gaussian kernel (RVM-G) instead of support vector machine with Gaussian kernel (SVM-G) to reduce the calculation cost in the use of 19 channels, common special patterns (CSPs) filtering, and adaptive collection (AC). Results show that using RVM-G instead of SVM-G reduced the ratio of the number of efficient vectors to the number of training data from 97% to 55%. At this time, the averaged classification accuracies (CAs) using SVM-G and RVM-G were, respectively, 77% and 79%, showing no degradation. However, the calculation cost was more than that using SVM-G because RVM-G necessitates high calculation costs for optimization. Furthermore, results show that CAs using RVM-G were weaker than SVM-G when the training data were few. Additionally, results showed that nonlinear classification was necessary for silent speech classification.

This paper serves as a beginning of feasibility study for speech prostheses using an imagined voice. Although classification for silent speech presents great potential, many feasibility problems remain.

1. Introduction

Daily life demands that we use verbal and non-verbal communication. However, severely handicapped individuals such as people with advanced amyotrophic lateral sclerosis (ALS), locked-in syndrome, or nasopharyngeal cancer have difficulty expressing their thoughts. Their caregivers also face difficulties when caring for patients. A brain–computer interface (BCI) has been developed to provide prosthetics for such individuals.

The anticipated benefits of the BCI are not merely confined to those individuals. They are expected to be useful for entertainment, personal communication, and game devices, and with preventive medical treatments for healthy individuals. When both healthy individuals and those with a disability use the same core technology, the demand shown by healthy people is expected to contribute to the welfare of handicapped individuals through improved production and reduced costs of assistive equipment. We seek to develop devices that are attractive for both handicapped and healthy people. The objective technology requires portability, high classification accuracy, and usability.

For these supporting prosthetics, many studies have been conducted using methods such as P300 speller [1], steady-state visual evoked potentials (SSVEP) speller [2], SSVEP cursor controller [3], and near infrared spectroscopy (NIRS) [4]. For the P300 and SSVEP spellers, users must gaze on the attempted word. With the SSVEP cursor controller, subjects must undergo training to move cursors using electroencephalography (EEG). In the method using hemodynamic response, e.g., NIRS, users must train in the mode of imagining calculations or imagining fast songs for detection. The methods described above necessitate training of skills that users have never developed in daily life.

The classification of silent speech is a simple method that requires no special training. Many silent speech interface studies have used electromyographic (EMG) signals [5], electromagnetic field measurements with implanted magnets [6], and ECoG signals detected using invasive electrodes [7]. The method using EMG signals requires electrodes mounted on the user's face or neck. The system is uncomfortable and fragile. The method using magnet implantation around a patient's mouth is effective, but it requires surgical operations. Severely paralyzed patients might accept

* Corresponding author. Tel.: +81 90 9969 1137.
E-mail address: macchan_ga@yahoo.co.jp (M. Matsumoto).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.10.023

surgical operations, but healthy individuals would not accept them. Moreover, any method using invasive electrodes necessitates surgical operations. The detection of silent speech by EEG is a good method in terms of portability and user-friendliness.

In the last paper, we classified silent speech of vowels using EEG, as measured using scalp electrodes, common spatial pattern (CSP) support vector machine with Gaussian kernel (SVM-G), and adaptive collection (AC) [8]. We demonstrated the potential of silent speech.

Regarding feasibility, some problems arose: The classification accuracies were insufficient when using the small number of electrodes and it needs multi-class classification and online processing. The SVMs are well known as a pairwise classifier. For online processing, classification speed is important and large calculation cost is problem, but SVM with Gaussian kernel (SVM-G) has hyper-parameters that must be optimized by cross validations.

The relevance vector machine (RVM) was proposed as a method to reduce the number of relevance vectors and optimize the hyper-parameters automatically [9,11].

In this study, we recorded EEGs of five healthy subjects when they imagined the vocalization of the Japanese vowels, /a/, /i/, /u/, /e/, and /o/, while they remained silent and immobilized. EEG recordings were band-pass filtered and divided into epochs. Liner RVM, RVM with Gaussian kernel, and SVM with Gaussian kernel were used to classify 2 of 5 vowels.

Our research started from vowels because Japanese syllables are based on five vowels, /a/, /i/, /u/, /e/, and /o/. Most Japanese syllables consist of one of the five vowels and a consonant. Therefore, we studied vowels at first.

2. Experiments

2.1. Subjects

The experiments involved five 21–24-year-old male participants (S1–S5). All subjects were native speakers of Japanese who were right-handed, as assessed by the Edinburgh Inventory [12]. No participant had any neurological disorder or noteworthy health problem. Experiments were conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects.

2.2. Experiments

Each subject was seated comfortably in an armchair with eyes closed to avoid the influence of visual activation. The subjects were coached beforehand and had rehearsed with actual movements a few times to ensure correct task execution. The subjects were then asked to imagine voice production for 1 s, while remaining silent and immobilized. The Japanese vowels, /a/, /i/, /u/, /e/, and /o/were imagined. Two tasks were conducted: the fixed order task and the random order task. The tasks used sound commands generated by a personal stereo device (Walkman NW-E053; Sony Corp.). Subjects, while hearing them through earphones, were instructed to perform the following tasks.

2.2.1. Fixed order task

The timings of onset for imagined vocalization in order were organized in the following manner.

Task: Subjects were instructed to imagine the voice production (imagined vocalization) of one of the vowels/a/,/i/,/u/,/e/, and/o/in fixed order for 1 s following 1 s for rest. The onset and ending of the imagined vocalization were signaled to the subjects using clicking sounds (Fig. 1(a)).

One trial stream consisted of 5 vowels $\times$ 13 times for about 2.2 min. Subjects 1–4 performed the experimental set four times.

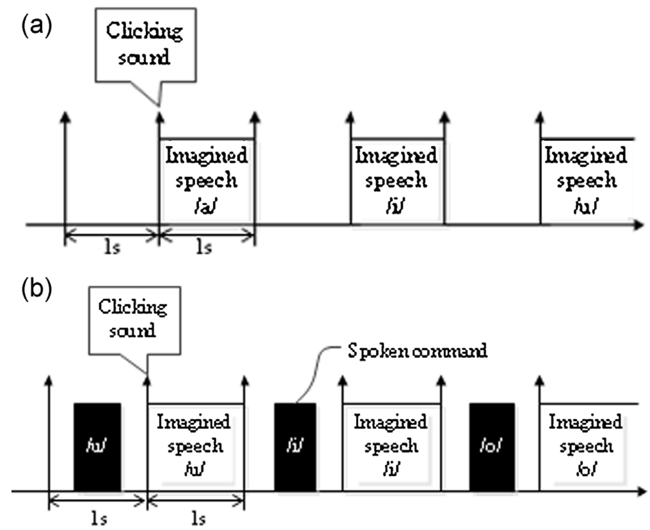

Fig. 1. Experimental protocols.

In all, 52 epochs were obtained for each vowel. Consequently, 260 trials were obtained for each subject. We designate this batch of data as 260 epochs. For subject 5, 65 epochs were obtained for each vowel from five experimental sets.

2.2.2. Random order task

The timings of onset for imagined vocalization were organized in the following manner.

Task: Subjects were instructed to imagine the voice production (imagined vocalization) of a vowel that was the same as the last spoken command for 1 s following 1 s for rest. The spoken command expressed one of the vowels, /a/, /i/, /u/, /e/, or /o/in randomized order. The onset and ending of the imagined vocalization were signaled using clicking sounds (Fig. 1(b)).

To avoid the influence of auditory evoked potentials, the interval between spoken commands and onset of the imagined speech was set to 200 ms or more. In all, 50, 45, 52, 52, and 65 epochs were obtained, respectively, for each vowel for subjects 1, 2, 3, 4, and 5. To clarify, in task /a/, for instance, subjects imagined speech production of /a/ for 1 s, while they remained silent and immobilized. The ways for other vowels are as the same as above.

2.3. Recording

EEG signals were recorded using an electroencephalograph (Neurofax EEG-1100; Nihon Kohden Corp.) and 128 channel Modular EEG-Recording Caps (Easycap GmbH) with a sampling rate of 1000 Hz. The recorded EEG data were zero-phase band-pass filtered at frequencies of 0.1–300 Hz to avoid anti-aliasing noise and to remove any low-frequency baseline shift.

We recorded brain waves using 111 electrodes in the previous study for subjects 1, 2, 3 and 4 and recoded brain waves using 19 electrodes of international 10–20 position for subject 5 [13]. For feasibility study, we reduced the number of electrodes from 111 to 19 for subjects 1, 2, 3, and 4. The remaining electrodes for calculation were therefore 19 electrodes. For reference, two electrodes were attached on the right and left ears. One electrode was set below an eye to detect unwanted eye movement and artifacts. However, no artifact rejection algorithm was used for this study.

2.4. Data processing

Data processing was performed using software (MATLAB; The MathWorks Inc., Natick, MA). Using a decimation filter with