Neurocomputing 80 (2012) 83–92

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Robust real-time identification of tongue movement commands from interferences

Khondaker A. Mamun [a], Michael Mace [b], Lalit Gupta [c], Carl A. Verschuur [a], Mark E. Lutman [a], Maria Stokes [d], Ravi Vaidyanathan [e,f,*], Shouyan Wang [a,**]

[a] Hearing and Balance Centre, Institute of Sound and Vibration Research, University of Southampton, UK
[b] Department of Mechanical Engineering, Faculty of Engineering, University of Bristol, UK
[c] Department of Electrical and Computer Engineering, Southern Illinois University, IL, USA
[d] Faculty of Health Sciences, University of Southampton, UK
[e] Department of Mechanical Engineering, Imperial College London, UK
[f] Department of Systems Engineering, US Naval Postgraduate School, USA

ARTICLE INFO

Available online 7 November 2011

Keywords:
Tongue-movement ear pressure signals
Wavelet packet transform
Bayesian classifier
Human machine interface

ABSTRACT

This study aimed to improve the accuracy and robustness of a real-time assistive human machine interface system by classifying between the controlled movements related tongue-movement ear pressure (TMEP) signals and the interfering signals. The controlled movement TMEP signals were collected during left, right, up, down, flicking and pushing tongue motions. The TMEP signals were processed and classified using detection, segmentation, feature extraction and classification. The segmented signals were decomposed into the time-scale domain using a wavelet packet transform. The variance of the wavelet packet coefficients and its ratio between low-to-high scales were defined as features and the intended tongue movement commands and interfering signals were classified using both a Bayesian and support vector machine (SVM) classifiers for comparison. The average classification accuracy for discriminating between the controlled movements and the interfering signals achieved 97.8% (Bayesian) and 98.5% (SVM). The classifiers were robust remaining at a similar performance level when generalised interferences from all subjects were used. It was shown that the Bayesian classifier performed better than the SVM in a real-time environment. The approach of combining the Bayesian classifier and the wavelet packet transform provides a robust and efficient method for a real-time assistive human machine interface based on tongue-movement ear pressure signals.

1. Introduction

A wide range of research has been conducted to develop various human-machine interfaces (HMI) based on human physiological signals for hands-free communication to assist physically impaired patients [1–6]. Specific hands-free communication and control devices are essential for an individual who has limited mobility or severe motor dysfunctions, for example due to spinal cord injury, congenital limb deformities or arthritis [5,7,8]. In spite of significant progress made in the development of techniques and devices for HMI systems, current products have not yet fully addressed patient-specific requirements and better interfaces between the patient and peripheral devices are still greatly needed [1,4,5,9]. Recently a novel hands-free communication concept based on tongue-movement ear pressure (TMEP) signals has been introduced [4,10,11]. Users express their intention by making impulsive actions of the tongue, which create unique acoustic pressure signals within the ear canal. These pressure signals can be recorded easily using a microphone earpiece positioned non-invasively within the ear canal [4]. The advantage of utilising the tongue is that it has an inherent capability for fine motor control, involving multiple degrees of freedom, as it has evolved to perform sophisticated motions during speech and mastication [1,4]. The system also has the additional benefits of being simple, cheap and non-invasive. Individuals with limited control of their limbs are able to use these prescribed tongue movements to communicate with computers and control assistive devices through the sensing of bio-acoustic pressure signals.

Previously, different types of tongue movements recorded from healthy subjects relating to the controlled (intended) actions

* Corresponding author at: Department of Mechanical Engineering, Imperial College London, London, SW7 2AZ, UK.
** Corresponding author at: Hearing and Balance Centre, Institute of Sound and Vibration Research, University Road, Southampton, SO17 1BJ, UK.
Tel.: +44 23 8059 2310; fax: +44 23 8059 3190.
E-mail addresses: r.vaidyanathan@imperial.ac.uk (R. Vaidyanathan), sy.wang@soton.ac.uk (S. Wang).
URL: http://www.swanglab.com (S. Wang).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.09.018

have been classified using a decision fusion algorithm [4]. The performance of the classifier reached an average of 97% correct accuracy using time domain features and large training sets. The performance of this classifier was shown to be better than three other strategies using time domain information, namely, the matched filter (86%), the parametric autoregressive (AR) Gaussian classifier (85.98%) and the nonlinear alignment classifier (96.27%) [4]. Moreover, to improve the classification performance a single channel independent component analysis (ICA) was used to isolate the critical components of TMEP signals, which are associated with four different tongue motions [10]. This method robustly extracted features and may be more useful when a higher number of movement commands are required or the signals are contaminated with noise. However, the higher computational load makes it unsuitable for real-time applications.

To explore the real-time implementation of a TMEP signal based assistive communication system, several challenges need to be addressed. One significant challenge is the ability of the system to classify TMEP signals in real environments under the influence of interference, including external noise from the surrounding environment (e.g. conversation, road noise), motion artifacts (e.g. head movements), internal noise or artifacts due to natural tongue movements (e.g. speech, mastication). Such interference problems are generally challenging in any human machine interface system. Superior performance of TMEP signal classification has been achieved in the data sets collected in controlled environments [4]. However, significant degradation was experienced when the TMEP signals were contaminated with noise [12,13]. A de-noising algorithm based on discrete wavelet thresholding was applied to improve the quality of signals [11]. Another challenge is that only a limited number of signals are available to train and calibrate the classifier in real environments [12,13]. On the other hand, the accuracy and robustness of a classification algorithm depends highly on its input and therefore optimal selection of its features is very important, especially in noisy environments.

As the TMEP signals of movement actions exhibit transient behaviour in the order of tens of milliseconds, the wavelet packet transform (WPT) should be able to reliably extract features in a multi-scale manner for the classification between movement commands and interferences. The WPT can capture localised time-frequency information of signals and has been implemented widely in signal analysis and modelling [14–18], with significant successful application in diverse fields such as signal detection, classification, compression, noise reduction and image processing [19–21]. To improve the classification accuracy in the presence of external interferences, the WPT was applied to extract features for classification of TMEP controlled actions [12,13,22]. Based on these WPT features, the classification performance has achieved a 97% recognition rate in a simulated noisy environment in comparison to poor performance (88%) with time domain features. In order to further improve the accuracy, reliability and robustness of the assistive HMI system based on TMEP signals, controlled movement related TMEP signals should be discriminated from a wide range of interfering signals that occur in daily life. These interfering signals can be categorised into non-controlled movement or interference related TMEP signals such as speech, swallowing, coughing, eating, drinking, and external artifacts such as the individual's heart beat, remote muscular activity, limb tremor and environmental sounds.

This study aimed to identify controlled movement related TMEP signals from a variety of interferences. The features were extracted using a WPT to capture the transient changes in the TMEP signals and were optimally selected according to statistical distributions of the wavelet packet coefficients so as to maximise the separability between movement commands and interferences. Two types of classifiers, a Bayesian and support vector machine (SVM), were implemented to perform the classification between two classes of commands and interferences. Their performance was evaluated in both offline and online conditions using both subject specific and generalised interference for training. This work has significantly improved the accuracy and robustness of both offline and online real-time assistive human machine interface systems based on TMEP signals.

2. Experimental paradigm and signal acquisition

2.1. Participants

Ten healthy subjects (6 males, 4 females) ranging from 20–45 years (30.7 ± 6.4; mean $\pm$ 1 SD) participated in the experiment. It is noted that within this subject group, five subjects (S6–S10) were well trained to perform the controlled TMEP actions whilst the remaining five subjects (S1–S5) had only half an hour practice prior to data collection. The experiment was approved by the ISVR Human Experimentation Safety and Ethics Committee of the University of Southampton. Participants gave their written informed consent before taking part in the study.

2.2. Experiment and signal recording

The oral cavity is connected to the ear via the Eustachian tube. Tongue movements cause pressure changes within the ear canal, which can be detected by a sensor. The sensor includes a shielded housing plug and an internal microphone. The microphone was inserted into the ear canal and connected to an amplifier. The pressure change was picked up by the microphone and digitised and stored in a computer similarly as in [4]. The distinct movement related actions can be differentiated from signatures of the recorded ear pressure signals.

In the present study the classification was performed between controlled movement commands and interference related TMEP signals. TMEP signals were recorded when subjects performed six types of controlled tongue movement: moving the tongue from the neutral position to the top/front centre of the roof of the mouth ('up'), touching the tongue to the bottom/front centre of the mouth ('down'), the front/right side of the mouth ('right'), the front/left side of the mouth ('left'), flicking the tongue up and down once ('flicking') and moving the tongue to the outside of the oral cavity in a straight manner with closed lips ('pushing'). TMEP signals during these six intended tongue actions were defined as controlled or intended movement related TMEP signals.

In contrast, non-controlled movement or interference related TMEP signals were collected while subjects were speaking, coughing, drinking or resting. The speech activity included utterances of words consisting of numbers from 0 to 9, and words 'start', 'stop', 'open', 'close', 'on' and 'off'. The drinking activity was to drink 15 ml of water from a glass, whilst the resting activity was recorded during normal relaxation. This set of words represents a wide range of tongue movement patterns.

Each subject was seated in a comfortable armchair with a recording microphone sensor inserted into the ear canal. Prior to the experiment, the selection of ear (left or right) to insert the earpiece was made by the participants according to individual preference. The signals were recorded using custom made software written in Microsoft C# running on a laptop computer.

A visual cue was presented on a computer screen to instruct the subjects to perform a specific tongue movement action. Subjects were instructed to move their tongue in the respective direction as much as possible, so as to perform each action correctly. The cues were represented by text as well as direction, via a moving circle on the screen. Before making each movement,