Medical Engineering & Physics 31 (2009) 1049–1055

Contents lists available at ScienceDirect

Medical Engineering & Physics

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

Swallow segmentation with artificial neural networks and multi-sensor fusion

Joon Lee [a,b,c], Catriona M. Steele [b,a,e], Tom Chau [a,d,*]

[a] Bloorview Research Institute, Toronto, Canada
[b] Toronto Rehabilitation Institute, Toronto, Canada
[c] Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada
[d] Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada
[e] Department of Speech-Language Pathology, University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Received 25 September 2008
Received in revised form 2 May 2009
Accepted 1 July 2009

PACS:
87.85.Ng

Keywords:
Deglutition
Swallow
Segmentation
Artificial neural network
Multi-sensor fusion

ABSTRACT

Swallow segmentation is a critical precursory step to the analysis of swallowing signal characteristics. In an effort to automatically segment swallows, we investigated artificial neural networks (ANN) with information from cervical dual-axis accelerometry, submental MMG, and nasal airflow. Our objectives were (1) to investigate the relationship between segmentation performance and the number of signal sources and (2) to identify the signals or signal combinations most useful for swallow segmentation. Signals were acquired from 17 healthy adults in both discrete and continuous swallowing tasks using five stimuli. Training and test feature vectors were constructed with variances from single or multiple signals, estimated within 200 ms moving windows with 50% overlap. Corresponding binary target labels (swallow or non-swallow) were derived by manual segmentation. A separate 3-layer ANN was trained for each participant–signal combination, and all possible signal combinations were investigated. As more signal sources were included, segmentation performance improved in terms of sensitivity, specificity, accuracy, and adjusted accuracy. The combination of all four signal sources achieved the highest mean accuracy and adjusted accuracy of 88.5% and 89.6%, respectively. A–P accelerometry proved to be the most discriminatory source, while the inclusion of MMG or nasal airflow resulted in the least performance improvement. These findings suggest that an ANN, multi-sensor fusion approach to segmentation is worthy of further investigation in swallowing studies.

1. Introduction

In swallowing research, several non-invasive signal modalities have been investigated for swallow assessment (e.g. pulse oximetry [1], neck vibrations [2,3], cervical auscultation [4], surface electromyography [5], or nasal airflow [6]). Such signal modalities are attractive due to their low cost, easy-to-attach sensors, and portability. Swallowing studies involving non-invasive signals have mainly focused on understanding the relationship between signal characteristics and swallowing function. However, to analyze the information buried in a signal about a particular swallow, a crucial precursory step is swallow segmentation. Although manual segmentation is always an option, automatic swallow segmentation is required for real-time swallow analysis or when dealing with massive volumes of data. In fact, automatic signal segmentation is a critical part of most computerized diagnostic systems, such as, for example, medical imaging equipment [7,8]. Automatic segmentation has also been investigated with electrocardiograms (ECG) (e.g. [9]), phonocardiograms (PCG) (e.g. [10]), and speech signals (e.g. [11]). Signal segmentation isolates specific segments of interest from a continuous stream of time series data. The segmented data are critical for informing diagnostic decision. The segments must capture the physiological phenomena under scrutiny, while minimizing contributions from other unwanted sources such as motion artifacts.

This study focuses on three non-invasive signal modalities, namely dual-axis accelerometry, submental mechanomyography (MMG), and nasal airflow, for automatic swallow segmentation. A brief background of these modalities is provided below, followed by a brief introduction about the two technical pillars of this study, namely, multi-sensor fusion and the artificial neural network (ANN).

The measurement of neck vibrations is motivated by cervical auscultation, which is based on the claim that normal and abnormal swallows exhibit audible differences [12]. Although a microphone can be used, the combination of an accelerometer and digital signal processing has been the recent focus of research and termed swallowing accelerometry [2,13–15]. One accelerometry study has shown a significant correlation between peak neck vibration and maximum hyolaryngeal excursion [3], which is a

* Corresponding author.
E-mail address: tom.chau@utoronto.ca (T. Chau).

1350-4533/$ – see front matter © 2009 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2009.07.001

vital biomechanical component of airway protection during swallowing [16]. Although most prior work in this area has examined single-axis accelerometry signals, it has recently been shown that the anterior–posterior and superior–inferior axes of dual-axis accelerometry contain distinct information about swallowing [13].

Deglutition (i.e. swallowing) is a sequence of well-coordinated muscle activations. More than 25 pairs of muscles in the oral cavity, pharynx, larynx, and esophagus participate in deglutition [17]. In particular, the submental muscles, including the mylohyoid, geniohyoid, and styloglossus muscles, are part of the group of muscles that contract first in deglutition, known as the swallowing leading complex [18]. Thus, the submental musculature has been extensively studied in relation to swallowing. Although electromyography (EMG) is the most common method of measuring muscle activity, MMG offers several advantages in swallowing studies such as tolerance to variations in electrode location and robustness to perspiration and food spillage [19,20].

Respiration and deglutition must be coordinated precisely in order to avoid airway invasion during swallowing [21], because the pharynx is a shared passageway for both air and food. The cessation of breathing during bolus transport is called swallowing apnea (SA), and the absence or presence as well as the timing of SA can disclose crucial information about airway protection [22,23]. Because the mouth is occupied with mastication (i.e. chewing) and bolus formulation during food intake, nasal respiration is the usual option for airflow monitoring in swallowing studies.

In a wide range of applications, multi-sensor fusion methodologies resulted in successful performance in the cases of both complementary and redundant data [24,25]. Although multi-sensor fusion has been applied to segmentation before, such efforts mainly focused on computer vision applications (e.g. [26,27]). Decisive advantages of multi-sensor fusion include reduced uncertainty, robustness to noise and measurement error in individual sensors, tolerance to single sensor failure, and resolved ambiguity [28].

ANN has been deployed in image segmentation problems (e.g. [29,30]). The ANN is a versatile nonlinear function approximator that can be utilized in either regression or classification. A 3-layer ANN is capable of mapping any input–output relationship, given sufficient hidden units and training cases as well as suitable nonlinear activation functions [31]. The input–output relationship can be learned automatically, and this is particularly helpful if the relationship is too complex to be described analytically. At least one study has investigated ANN classifiers based on multiple bio-sensors for emotion recognition [32].

To the best of our knowledge, a multi-sensor fusion ANN has never been investigated for swallow segmentation. Furthermore, the combination of dual-axis accelerometry, submental MMG, and nasal airflow signals is novel in the field of swallowing research.

2. Objectives

The two primary objectives of this study were:

- to investigate the relationship between swallow segmentation performance and the number of employed signal sources, and
- to determine the signals or signal combinations which yield the most accurate swallow segmentation.

3. Methods

3.1. Signal acquisition

Signals were acquired from 17 (8 male) healthy adults with no history of dysphagia (i.e., swallowing disorder) or neurological impairments. The mean age was 46.9 ± 23.8 years. Each participant's swallowing health was confirmed via a standardized oral mechanism examination and a water swallow screening test conducted by a registered speech-language pathologist. This study was approved by the ethics committees at the Toronto Rehabilitation Institute, Bloorview Kids Rehab, and University of Toronto. All participants gave informed consent.

Each participant completed 19 swallowing sequences, each comprising four swallows, in a comfortably seated position. Swallowing sequences were either discrete or continuous. In the discrete swallowing task, participants were instructed to take four separate sips from a cup, removing the cup from the lips between consecutive sips. In the continuous swallowing task, participants were required to perform four consecutive swallows in a continuous manner, without taking the cup away from the lips between sips. In addition, participants were asked to initiate each swallowing sequence upon hearing an audible cue, which was generated 5 s after the beginning of signal acquisition.

The following five stimuli were presented to each participant:

- Water
- 40% weight per volume barium suspension (prepared with water and Liquid Polibar ™ barium from E-Z-EM)
- Nectar-thick apple juice (RESOURCE® by Novartis Nutrition)
- Honey-thick apple juice (RESOURCE® by Novartis Nutrition)
- Spoon-thick apple juice (prepared using honey-thick RESOURCE® and Quick-Thick powder, both by Novartis Nutrition).

Participants were asked to perform both the discrete and continuous swallowing tasks for all stimuli except for the spoon-thick apple juice, which was self-administered by spoon, necessitating a discrete method of swallowing. Each stimulus–task combination was repeated twice except the water-discrete combination, which was repeated 3 times.

Four signal sources were targeted: submental MMG, nasal airflow, and two axes of cervical vibrations. For the two vibration signals, a dual-axis accelerometer (ADXL322, Analog Devices) was placed on the neck just below the thyroid cartilage with the axes oriented in the anterior–posterior (A–P) and superior–inferior (S–I) directions. Sensor attachment was accomplished with a double-sided electrode collar (650455, VIASYS Healthcare). The MMG signal was acquired via the microphone within the novel sensor developed by Silva and Chau [33]. The MMG sensor was attached with tape and an elastic, custom-made fastener that hung over the ears. Finally, the nasal airflow signal was obtained with a nasal cannula (Pro-Flow Cannulas Model 1259, Grass Technologies) which was connected to a pressure transducer (PTAFLITE, Grass Technologies). Fig. 1 pictorially describes the locations of the sensors with respect to related anatomical structures, as well as the orientation of the two accelerometry axes.

All four signal channels were sampled in synchrony at a sampling frequency of 10 kHz using a custom LabVIEW application. Each signal was subsequently passed through a pre-amplifier with a bandpass filter (Model P55, Grass Technologies), the cutoff frequencies of which were set at 0.1 Hz and 3 kHz. Amplification was set at 10 times for the two accelerometry signals and 10,000 times for the nasal airflow signal. No amplification was needed for the MMG signal.

3.2. Pre-processing

The sampling frequency of 10 kHz was initially selected because the accelerometry signals had the potential to possess high spectral content. However, our spectral analysis revealed that most signal power lies below 100 Hz for the accelerometry signals [13]. In addition, the MMG signal bandwidth typically ranges from 10 to 40 Hz [34] and respiration at rest occurs only 10–12 times per minute in