Available online at www.sciencedirect.com
ScienceDirect
Medical Engineering & Physics 30 (2008) 419–425
Medical Engineering & Physics
www.elsevier.com/locate/medengphy

Development of a (silent) speech recognition system for patients following laryngectomy

M.J. Fagan [a,*], S.R. Ell [b], J.M. Gilbert [a], E. Sarrazin [a], P.M. Chapman [c]

[a] *Department of Engineering, University of Hull, UK*
[b] *Department of Otolaryngology, Hull Royal Infirmary, Hull and East Yorkshire Hospitals NHS Trust, UK*
[c] *Department of Computer Science, University of Hull, UK*

Received 20 November 2006; received in revised form 2 May 2007; accepted 3 May 2007

Abstract

Surgical voice restoration post-laryngectomy has a number of limitations and drawbacks. The present gold standard involves the use of a tracheo-oesophageal fistula (TOF) valve to divert air from the lungs into the throat, which vibrates, and from this, speech can be formed. Not all patients can use these valves and those who do are susceptible to complications associated with valve failure. Thus there is still a place for other voice restoration options.

With advances in electronic miniaturization and portable computing power a computing-intensive solution has been investigated. Magnets were placed on the lips, teeth and tongue of a volunteer causing a change in the surrounding magnetic field when the individual mouthed words. These changes were detected by 6 dual axis magnetic sensors, which were incorporated into a pair of special glasses. The resulting signals were compared to training data recorded previously by means of a dynamic time warping algorithm using dynamic programming. When compared to a small vocabulary database, the patterns were found to be recognised with an accuracy of 97% for words and 94% for phonemes. On this basis we plan to develop a speech system for patients who have lost laryngeal function.

Keywords: Speech recognition; Rehabilitation; Laryngectomy; Magnetic sensor; Speech system

1. Introduction

Patients with laryngeal cancer, whose larynx must be removed, inevitably lose their voice. Also, as a result of surgery, the viscera involved in swallowing and breathing are separated so that the patient must breathe through their neck via a permanent tracheostomy. The three main methods used currently to restore vocal function may encounter a number of problems and limitations. Sound can be created by swallowing air and belching, forming the sound into words. This is known as 'oesophageal speech' and is difficult to learn, and fluent speech is impossible. Vibrating the soft tissues of the throat by an electrolarynx creates sound, which can be articulated into speech, but the voice is monotonic, 'Dalek-like', and can be difficult to understand. The current 'gold-standard' method is to use a small silicone tracheo-oesophageal fistula speech valve that connects the trachea and the oesophagus [1]. Air, powered by the lungs, is diverted through the fistula into the throat which vibrates, and this is formed into speech. However, although these valves work very well initially, they rapidly become colonised by biofilm in many patients and fail after an average of only 3–4 months [2–5]. Various modifications have been tried over the years to discourage biofilm growth (e.g. [6–8]), but to date none of these approaches appears to provide a long-term solution to this problem.

Thus there is a need for a fundamental improvement in the current methods for the restoration of speech after laryngectomy. Digital (voiced) speech recognition systems have been the subject of research for a number of years, based on measurement of sound emitted by the speaker [9] and a variety of

* Corresponding author at: Centre for Medical Engineering and Technology, Department of Engineering, University of Hull, Hull HU6 7RX, UK. Tel.: +44 1482 465058; fax: +44 1482 466664.
E-mail address: m.j.fagan@hull.ac.uk (M.J. Fagan).

1350-4533/$ – see front matter © 2007 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2007.05.003

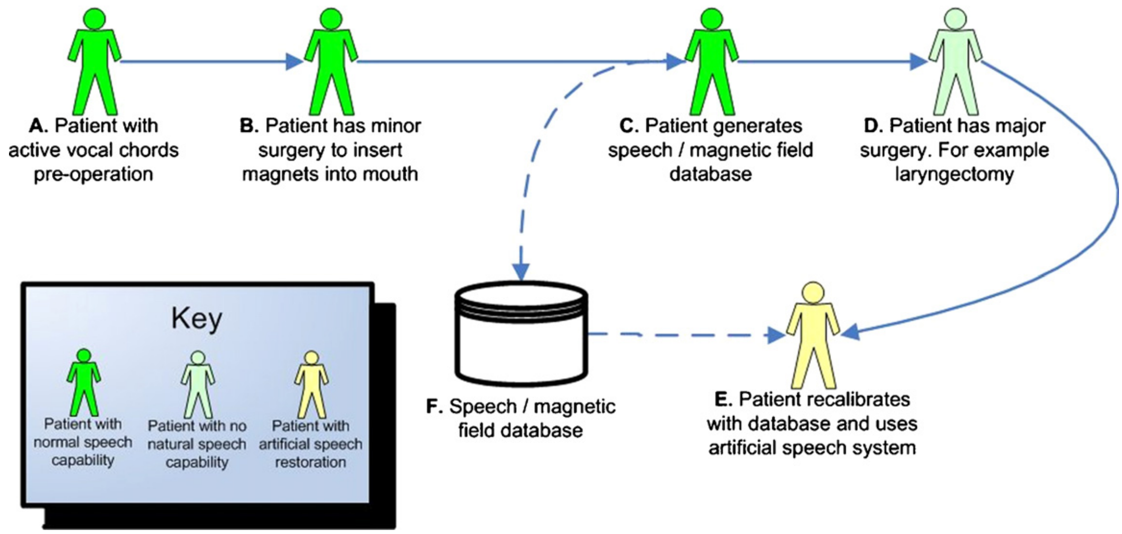

Fig. 1. Overview of patient status pre- and post-laryngectomy.

methods exist for identifying the output (e.g. [10–12]). While these techniques give good recognition rates (typically 90% for continuous speech), they are not perfect, particularly in noisy environments and so consideration has been given to the augmentation of acoustic signals by other measurements [13–16].

A new approach has recently been proposed by the authors [17] that tracks how an individual's mouth and tongue move when they 'speak' (i.e. mouth the words). Analysis of this information allows a computer to decide what the individual had 'said' and then plays back those words. Eventually this computer generated speech could be in the individual's original voice. If developed, such a system could restore speech to patients who have lost their voice as a result of throat cancer, trauma, destructive throat infections or damage to the laryngeal nerves.

The essence of the proposed system is that by monitoring the motion of the vocal apparatus, it is possible to determine the phonemes and words that an individual wishes to produce. A number of miniature magnets implanted into the appropriate parts of the patient's mouth (e.g. lips, tongue and teeth) will result in a variation in the magnetic field surrounding the mouth during 'speech'. By monitoring these variations in magnetic field and comparing these variations with a database of pre-recorded signals, it is proposed that the best matching word or phoneme may be identified. It is envisaged that the implants would be small enough that they would not be apparent when observing the patient and would not, affect the movement of the mouth, either for sound generation or for eating, etc. The implants used in the ideal system are expected to be 1–2 mm in size and coated with a biocompatible material, e.g. silicone. These may be hidden in the teeth of dentures, or implanted in, or bonded to, the posterior surface of the incisors using standard dental techniques, so that they are hidden from view. Implants may be placed into the lips and tongue by an injection technique under local anaesthetic. It is envisaged that the motion sensing system would be incorporated into the patient's normal attire, for example, in a pair of glasses or a necklace.

Fig. 1 provides an overview of a patient's status pre- and post-laryngectomy for laryngeal carcinoma. At the time of diagnosis, our patient will be able to speak (Fig. 1A), even though his/her voice may have been affected by the disease. Once it has been established that the curative procedure is a laryngectomy the patient may then enter the speech rehabilitation program. After radiographic imaging has been obtained for surgical planning, the patient undergoes minor surgery (Fig. 1B) to implant magnets as described previously. The patient will then undergo a sequence of clearly defined recording procedures (Fig. 1C) in order to generate a complete speech/magnetic field database unique to the patient (Fig. 1F). This process should not delay the patient's progress to curative surgery and would form part of the usual pre-operative work up. After this database has been successfully constructed and tested, the patient will undergo a laryngectomy (removing the patient's ability to speak, Fig. 1D). When the patient has recovered from this operation, they should be able to recalibrate (Fig. 1E) and use the speech system immediately conversing in their original pre-laryngectomy 'normal' voice.

At present, the research activity described here is concerned with the identification of the words and/or phonemes mouthed by the patient, with the aim of establishing whether such a system is feasible. There are a number of sophisticated speech generation systems already available that can produce voiced output from word or phoneme data; therefore, such a system will be used to recreate the individual's speech in this application.