journal homepage: www.intl.elsevierhealth.com/journals/ijmi

Impact of noise and other factors on speech recognition in anaesthesia

*Alexandre Alapetite**

Risø National Laboratory, Systems Analysis Department, SYS-110, P.O. Box 49, Frederiksborgvej 399, DK-4000 Roskilde, Denmark

ARTICLE INFO

Article history:
Received 19 July 2006
Received in revised form
8 September 2006
Accepted 20 November 2006

Keywords:
Anaesthesia
Electronic medical records
Voice recognition
Speech recognition
Noise
Error rates

ABSTRACT

Introduction: Speech recognition is currently being deployed in medical and anaesthesia applications. This article is part of a project to investigate and further develop a prototype of a speech-input interface in Danish for an electronic anaesthesia patient record, to be used in real time during operations.

Objective: The aim of the experiment is to evaluate the relative impact of several factors affecting speech recognition when used in operating rooms, such as the type or loudness of background noises, type of microphone, type of recognition mode (free speech versus command mode), and type of training.

Methods: Eight volunteers read aloud a total of about 3600 typical short anaesthesia comments to be transcribed by a continuous speech recognition system. Background noises were collected in an operating room and reproduced. A regression analysis and descriptive statistics were done to evaluate the relative effect of various factors.

Results: Some factors have a major impact, such as the words to be recognised, the type of recognition and participants. The type of microphone is especially significant when combined with the type of noise. While loud noises in the operating room can have a predominant effect, recognition rates for common noises (e.g. ventilation, alarms) are only slightly below rates obtained in a quiet environment. Finally, a redundant architecture succeeds in improving the reliability of the recognitions.

Conclusion: This study removes some uncertainties regarding the feasibility of introducing speech recognition for anaesthesia records during operations, and provides an overview of the interaction of several parameters that are traditionally studied separately.

1. Introduction

This paper reports some preliminary experiment about the effects of various background noises in the hospital operating room (OR) environment on speech recognition. The envisaged audio interface would supplement existing electronic anaesthesia record systems with voice input facilities during the operation. This work is part of a project seeking to investigate [1] and further develop a prototype of such a system in Danish.

During the experiment, eight participants read aloud a corpus of typical anaesthesia comments to be transcribed by a continuous speech recognition system. The main goal of the study was to measure the respective impact on the recognition rate of various parameters, namely the type or loudness of background noises, the type of microphone (headset or handheld) and the type of recognition mode (free speech versus command mode). Additional parameters were also investigated, including the type of training (with or without

* Tel.: +45 46 77 51 82; fax: +45 46 77 51 99.
E-mail address: alexandre.alapetite@risoe.dk.

1386-5056/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2006.11.007

background noise) and the gender of the participants. A logistic regression analysis was done to estimate the significance of each of the evaluated parameters.

As far as the author knows, this is the first study reporting the effect of background noises on speech recognition in Danish and the first to compare the relative impact of the above parameters, all known to separately affect speech recognition, but not yet studied in parallel. Finally, a redundant cross-matching high level architecture was tested and shown to improve recognition rates.

2. Methodology

2.1. Preparatory work

To ensure the reproducibility of the background noises, it was decided to carry out the experiment in a laboratory rather than in the real-life context of a hospital OR. Some background noises were recorded in an OR (Herlev University Hospital of Copenhagen) during real anaesthesias with surgery and X-rays, using a multi-directional microphone placed in the proximity of the anaesthesiologist. Simultaneously, an integrating sound level meter (from Brüel & Kjær, model 2225) was used to measure the peak level and fixed level in dB(A) of various sounds. The 60 s L_{eq}[1] in dB(A) was also calculated for the background noise made by the room ventilation. The measurements have been made from the place where the anaesthesiologist is usually standing, and by pointing the sound level meter toward the various sound sources.

The collected sound files were edited and samples selected. Samples of the same type of noise were concatenated to create longer sequences with the same type of noise. The nine "background noises" were:

(1) "Silence": the laboratory background noise ∼32 dB(A);
(2) "Ventilation1": the constant background noise in the OR, air conditioning and pulse beeps, 48–63 dB(A), slow measure 60 dB(A), peak 70 dB(A);
(3) "Alarms": a set of classic anaesthesia alarms using various tones, 57–68 dB(A), peak 80 dB(A);
(4) "Scratch": velcro noise when opening anti X-ray suites 82 dB(A);
(5) "Aspiration": suction of saliva in the patient's mouth 65 dB(A);
(6) "Discussion": female voices, discussions between the surgeon 60 dB(A) and the nurse 70 dB(A);
(7) "Metal": various metallic clinks, 58–82 dB(A), peak 97 dB(A), this is the noise with the sharpest peaks;
(8) "Ventilation2": Same as "Ventilation1" but 10 dB(A) louder, giving 61–73 dB(A);
(9) "Ventilation3": Same as "Ventilation1" but 20 dB(A) louder, giving 71–83 dB(A), slow measure 75 dB(A).

2.1.1. Reproducing sounds

Samples were reproduced with a computer plugged to an audio amplifier (Sony STR-GX290) with two loudspeakers (Jamo Compact 1000, 65 Hz to 20 kHz, 90–120 W), positioned 1.5 m apart and pointing toward participants about 2 m away. This is similar to the distance from the anaesthesiologist to the noise sources in a real OR. The samples were played in a loop as long as needed.

In order to replay the samples at the appropriate volume, the sound level meter was used again from the position where the participants would be sitting, pointing in the direction of the loudspeakers. The replay volume was adjusted to match as closely as possible the measured values in dB(A).

2.2. Experiment

2.2.1. Speech recognition software

The lab experiment was made with the speech recognition system Philips[2] SpeechMagic 5.1.529 SP3 (March 2003) and SpeechMagic InterActive (January 2005), with a package for the Danish language (400.101, 2001) and a "ConText" for medical dictation in Danish (MultiMed Danish 510.011, 2004) from Philips in collaboration with the Danish company Max Manus.[3] The speech recognition workflow is the same as detailed in [2].

For voice dictation in free speech mode, or "natural language", SpeechMagic is integrated with Microsoft Word 2003. At the time of writing this article, a similar speech recognition system was already in use and under further deployment at Vejle Hospital (Denmark), for pre- and post-operative tasks, but not during operations [1]. With this system it is possible to record what is being said and to submit the WAV file for recognition afterwards; this was the process used for this experiment.

For voice commands, or "constrained language", SpeechMagic InterActive uses grammars [3] describing the set of possible commands. The grammar must contain the phonetic transcription of the terms used, for which the "Phonetic Transcriber component" can help.

Philips Speech Magic is now available in various languages, is no longer batch only (i.e., documents can be navigated and corrected while dictated) and has an interactive mode combining free text and command mode.

2.2.2. Hardware

Two similar laptop computers were used, running identical software. USB connections were chosen for microphones, since the noise added when using the analog mini-jack input to the sound card of the laptop computers noticeably reduced speech recognition accuracy. Two different microphones were employed, one per laptop, in order to evaluate the impact of these on the speech recognition quality. On PC#1, the microphone was a Philips SpeechMike Classic USB 6264[4] (Mic#1). This was the recommended model for the Philips SpeechMagic system. It is a Dictaphone-like device, held in one hand about 15 cm from the mouth. On PC#2, a headset microphone was used (Mic#2, ∼2.5 cm from the mouth),

[1] L_{eq}: equivalent continuous sound pressure.

[2] [http://www.speechrecognition.philips.com].
[3] [http://www.maxmanus.dk].
[4] [http://www.dictation.philips.com/index.php?id=1470].