Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 36 (2016) 110–121

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

Preprocessing for elderly speech recognition of smart devices☆

Soonil Kwon*, Sung-Jae Kim, Joon Yeon Choeh

Interaction Technology Laboratory, Department of Digital Contents, Sejong University, 98 Gunja-Dong, Gwangjin-Gu, Seoul 143-747, Republic of Korea

Received 13 October 2014; received in revised form 8 July 2015; accepted 4 September 2015
Available online 12 September 2015

Abstract

Due to the increasing aging population in modern society and to the proliferation of smart devices, there is a need to enhance speech recognition among smart devices in order to make information easily accessible to the elderly as it is to the younger population. In general, speech recognition systems are optimized to an average adult's voice and tend to exhibit a lower accuracy rate when recognizing an elderly person's voice, due to the effects of speech articulation and speaking style. Additional costs are bound to be incurred when adding modifications to current speech recognitions systems for better speech recognition among elderly users. Thus, using a preprocessing application on a smart device can not only deliver better speech recognition but also substantially reduce any added costs. Audio samples of 50 words uttered by 80 elderly and young adults were collected and comparatively analyzed. The speech patterns of the elderly have a slower speech rate with longer inter-syllabic silence length and slightly lower speech intelligibility. The speech recognition rate for elderly adults could be improved by means of increasing the speech rate, adding a 1.5% increase in accuracy, eliminating silence periods, adding another 4.2% increase in accuracy, and boosting the energy of the formant frequency bands for a 6% boost in accuracy. After all the preprocessing, a 12% increase in the accuracy of elderly speech recognition was achieved. Through this study, we show that speech recognition of elderly voices can be improved through modifying specific aspects of differences in speech articulation and speaking style. In the future, we will conduct studies on methods that can precisely measure and adjust speech rate and find additional factors that impact intelligibility.

Keywords: Elderly voice interface; Speech recognition; Aging society

1. Introduction

Demographics among developed nations around the world are shifting towards an aging population; for instance, the proportion of the elderly population in South Korea has reached unprecedented levels (Korea National Statistical Office, 2006). As the market penetration of information technology has spread widely, smart devices are not only used by young people for prolonged periods but also by the elderly with high usage indoors and outdoors during the entire day. Advanced interface methods supported by smart devices are largely conventional interfaces that depend on touch, voice, and motion commands. But touch-sensitive technology and finger-recognition motions used on smart devices

☆ This paper has been recommended for acceptance by R.K. Moore.

* Corresponding author. Tel.: +82 2 3408 3847.
E-mail addresses: skwon@sejong.edu (S. Kwon), kkskyo0@naver.com (S.-J. Kim), zoon@sejong.ac.kr (J.Y. Choeh).

http://dx.doi.org/10.1016/j.csl.2015.09.002
0885-2308/

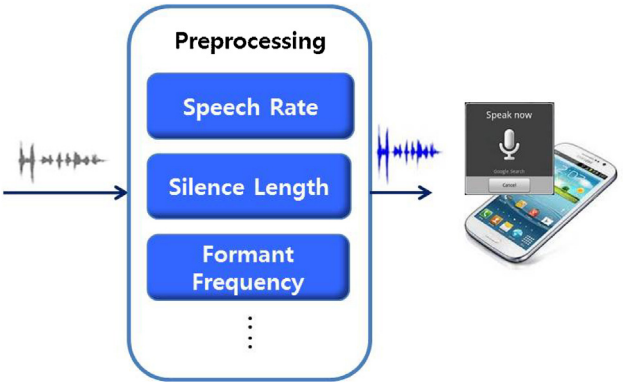

Fig. 1. Illustration of the proposed preprocessing method for elderly speech recognition.

are difficult to use among the elderly due to deterioration of eyesight and sense of touch. We therefore deemed that a speech-recognition interface in a smart device would increase its ease-of-use for the elderly. This interface would especially be utilized by senior citizens in emergency situations including a traumatic event that renders them physically limited.

Voice interfaces embedded in existing smart devices use speech recognition methods that are optimized to the speech and speaking patterns of average young and middle-aged adults and therefore, show a decline in performance when there is a deviation in the speech rate. Furthermore, tongue thickness, range of movement and duration of movement decreases as humans reach old age (Bennett et al., 2007) and due to this, feedback becomes slower when articulating speech sounds, in which it can partially affect articulation functions (Sonies, 1987). In other words, among the elderly, speech rate becomes slower, silence lengths are longer and speech becomes less precise (Kahane, 1981).

This study aims to investigate the factors that impair speaking ability among the elderly by comparing speech patterns between the elderly and young adults while also identifying which areas of speech that can be normalized or adapted for improving speech recognition. After the analysis of elderly speech, a preprocessing method was conducted to modify voice signals from elderly adults, which were then tested to determine if accuracy was improved in speech recognition without any modification of an automatic speech recognition system installed in an existing smart phone (Fig. 1). Test results showed that voice signals from the elderly increased the error rates among existing speech recognition systems, but accuracy showed marked improvement when the speech rate, silence lengths, and formant of the voice signals were preprocessed. The study results show promise in offering better accessibility to smart devices among the elderly and the speech-impaired who are left behind in the information age due to the lack of modifications in automatic speech recognition systems installed in existing smart phones.

In Section 2, speech patterns of the elderly and related studies on speech recognition are summarized, and in Section 3, the methods used to analyze and measure characteristics of speech patterns among the elderly are described. Section 4 goes into the comparative analysis of speech patterns between young adults and the elderly in order to identify differences in speech patterns. The speech signals are subsequently preprocessed for these differences and tested with existing automatic speech recognition systems installed on a smart device, and the results are analyzed. The conclusion of this paper is given in Section 5.

2. Previous research

In this section, we provide a brief review of studies on elderly speech based on prior technological research in the field of automatic speech recognition as well as research in fields such as applied linguistics and biomechanics. When humans reach old age, their eyesight, cognitive processes, sensory abilities, and physical functions that allow them to communicate verbally decline in comparison to younger people. The occipital lobe, which controls muscle movement and the lungs, shows a marked decline while the mucous membrane in the vocal cords becomes thinner and keratinized. In addition, the thyroid cartilage in men over 65 becomes nearly completely ossified, while, in women, ossification is limited to the lower part of the cartilage (Lee, 2011). Due to these aging factors, people over the age of 65 exhibited a significantly slower articulation speed and speech rates compared to that of the younger generation, and also have a