Automatic speech recognition for edentulous speakers with insufficient dentures

Tobias Bocklet (tobias.bocklet@informatik.uni-erlangen.de), Florian Stelzle, Tino Haderlein, Elmar Nöth

Introduction

A complete loss of teeth can cause persistent speech disorders by altering dental articulation areas. Dental rehabilitation by complete removable dentures is a state-of-the-art approach to improve functional aspects of the oral cavity of edentulous patients. In this work we introduce a dataset of 13 edentulous patients that have been recorded with and without complete dentures in situ while they spoke the text „Der Nordwind und die Sonne“ (NWS). These patients have been rated an insufficient of their dentures, so that additional, sufficient, dentures and additional speech recordings have been prepared. We performed automatic speech recognition (ASR) experiments w.r.t. these three different subsets, i.e., without any dentures, with sufficient dentures and with insufficient dentures. The performance of the ASR increases when dentures are worn. Sufficient dentures allow an additional improvement compared to insufficient dentures.

Material & Methods

The original version of the dataset contains 28 edentulous speakers. The average age was 64±10 years. 13 out of these patients have been rated an insufficient fit of the dentures and new dentures have been produced for these patients. In this work we used three recordings of these 13 speakers: without any dentures, with sufficient dentures, with insufficient dentures. Each recording contains audio data of the NWS-text.

Based on these 39 recordings (13 x 3) we performed ASR experiments in order to evaluate recognition differences between each of the three kinds of recordings. The ASR system itself is based on semicontinuous Hidden Markov Models (HMMs). Phones are modeled in a variable context, the so-called polyphones. As recognition measure we calculate the word accuracy (WA)

$$WA = [1 - (\#SUB + \#DEL + \#INS / \#ALL] \times 100$$

w.r.t. to the original words of the NWS text. Here, #SUB denotes the number of substituted words, #DEL the number of deleted words and #INS the number of wrongly inserted words. #ALL denotes the number of all words, in case of the NWS text #ALL is 71.

Results

The results are summarized in Table 1. The ASR system achieved a mean WA of 60.1% when no dentures are worn, 64.4% with insufficient dentures and 70.9% with sufficient dentures.

Dataset	mean WA	min WA	max WA
Without	60.1 %	39.5 %	77.8 %
Insufficient	64.4 %	45.4 %	80.6 %
Sufficient	70.9 %	57.5 %	80.6 %

Table 1: Word accuracy results for the three different subsets

Discussion

The ASR system achieved the lowest WA when the speakers did not wear any dentures at all. With insufficient dentures the WA improved by 7.3%. Comparing teethless recordings with recordings of sufficient dentures the WA improved by 18.2%. Wearing sufficient dentures improves the performance of an ASR system. That means we can measure the acoustic differences by means of speech recognition. In a next step we try to employ this result in order to automatically detect if dentures have an insufficient fit or not.