Biomedizinische Technik / Biomedical Engineering Vol. 52 (2007)
41. Jahrestagung der Deutschen Gesellschaft für Biomedizinische Technik (DGBMT)
im VDE 26.-29.09.2007
de Gruyter, Berlin, New York, 2007, ISSN 0939-4990
No pagination

PEAKS – a Platform for Evaluation and Analysis of all Kinds of Speech Disorders

Dipl.-Inf. Andreas Maier, Dipl.-Inf Tino Haderlein, PD Dr. med. Maria Schuster, Abteilung für Phoniatrie und Pädaudiologie des Universitätsklinikum Erlangen, Bohlenplatz 21, 91054 Erlangen
PD Dr.-Ing. Elmar Nöth, Lehrstuhl für Mustererkennung, Martensstr. 3, 91058 Erlangen

Abstract

We present a computer system for the automatic evaluation of speech disorders. The system can be accessed via internet or public telephone. A text or pictures to be named are presented on the computer screen. Speech data is then analyzed by automatic speech recognition technique. The measurements do not require experts' knowledge but yield results on experts' level. This is shown for the global outcome parameter of disordered speech – the intelligibility.

1 Introduction

In this paper, we present PEAKS (**P**latform for **E**valuation and **A**nalysis of all **K**inds of **S**peech disorders), a new recording and analysis environment for the automatic evaluation of speech disorders for medical purposes. Communication disorders are frequent but their diagnostic methods lack standardized and objectified evaluations. Until now, diagnostics still rely on perceptual assessments by human experts with only moderate reliability and accuracy. The new system ameliorates the medical diagnostic procedure and avoids time-consuming perceptual assessment.

Its validity for intelligibility as the global parameter for the speech outcome was tested on different speech disorders (children with cleft lip and palate and patients with cancer in the oral cavity). Both corpora contained a wide range in intelligibility.

2 Materials and Methods

2.1 PEAKS

PEAKS is a client-server system implemented in Java. The client-applet runs in any web-browser and is thus platform-independent. All transfer between the client and the server is encrypted using SSL sockets and ensures the privacy of the patients' data.

The system can easily be accessed via internet, i.e. the system does not require either special hardware or software except for a standard PC with internet access, a soundcard, and a microphone nor expert's knowledge. The patient performs a standardized test for speech assessment presented on the screen divided into smaller segments. Acoustic data is then automatically evaluated w.r.t. different aspects such as intelligibility in half of real time. For the analysis we use an automatic speech recognition (ASR) system. The outcome of the ASR system is the word recognition rate. The result numerically describes the degree of the disorder on a continuous scale.

2.2 Automatic Speech Recognition System

For the objective measurement of the intelligibility of pathologic speech, we use a hidden Markov model (HMM) based ASR system. It is a state-of-the-art word recognition system developed at the Chair of Pattern Recognition (Lehrstuhl für Mustererkennung) of the University of Erlangen-Nuremberg. In this study, the latest version as described in detail in [1,2] was used.

A commercial version of this recognizer is used in high-end telephone-based conversational dialogue systems by Sympalog (www.sympalog.com), a spin-off company of the Chair of Pattern Recognition.

We use a standard feature vector as input to the HMM recognizer: 11 Mel-Frequency Cepstrum Coefficients (MFCCs) and the energy of the signal. In addition, 12 delta coefficients are computed.

The recognition is performed with semi-continuous Hidden Markov Models. The codebook contains 500 full covariance Gaussian densities. The elementary recognition units are polyphones [3], a generalization of triphones. Polyphones use phones in a context as large as possible which can still statistically be modeled. The HMMs for the polyphones have three to four states. The polyphones were constructed for each sequence of phones which appeared more than 50 times in the training set.

2.3 Speech Data

Recordings of 46 adults with speech disorder due to oral carcinomas, mean age 59.8 ± 10.1 years, form the first corpus. They read the fable "The North Wind and the Sun", which is a phonetically rich text with 108 words (71 disjoint).

The second patient database consists of recordings of 34 children and adolescents with cleft lip and palate. The patients of this database were between 3.5 and 18.7 years old (μ=8.6 ± 3.5) and spoke a standard test for children's speech.

All patients were recorded with a close-talking microphone at a sampling frequency of 16 kHz and quantized with 16 bit.

2.4 Subjective Evaluation

In order to prove the validity of the new system we compared its results to the perceptual evaluations of the intelligibility by human speech experts. The experts rated each segment of a test on a Likert scale between 1 (very good) and 5 (very bad). So a floating point value was computed for each patient to represent his intelligibility. The datasets were rated by a panel of 4 experts.

3 Results

The recordings showed a wide range of intelligibility (Adults WR 49 ± 19; Children WR 48 ± 18). Subjective speech evaluation showed good consistency. The lowest correlation value between a rater and the mean of the other raters is 0.81, the highest 0.95 (see Table 1). The results for the correlations of the WR and the subjective speech evaluation are shown in Table 2.

Rater	Correlation to the mean of the other raters Adults	Correlation to the mean of the other raters Children
Rater 1	0.91	0.92
Rater 2	0.81	0.95
Rater 3	0.91	0.90
Rater 4	0.95	0.93

Table 1: Agreement between the different raters

When compared to the average of the raters, the WR for the recognizer has a high correlation for both patient groups. The coefficients are negative because high recognition rates represent "good" speech with a low score number and vice versa.

Speech Recognizer	Correlation to the mean of the raters
Adults speech	-0.92
Children's speech	-0.90

Table 2: Agreement between the human raters and the speech recognition system

4 Discussion

The method presented evaluates the intelligibility as good as a panel of experts. It is suitable for the evaluation of speech for clinical and scientific purposes. The recording environment is highly accepted by the medical staff. One major reason is that there is no installation cost, since practically all examination rooms already have access to a PC with internet. The analyses can be performed by non-experts in half of real time, minimizing time-consuming experts' evaluations for the assessment of speech. Of course, it does not replace the interpretation of the results according to any other diagnostic tool. Currently, we are expanding the method to other German clinics.

In conclusion our evaluation system provides an easy to apply, time effective, instrumental, and objective evaluation for disordered speech.

5 References

[1] F. Gallwitz, Integrated Stochastic Models for Spontaneous Speech Recognition, Volume 6 of Studien zur Mustererkennung, Logos Verlag, Berlin, 2002.

[2] G. Stemmer, Modelling variability in speech recognition. Logos-Verlag, Berlin, 2005

[3] E. G. Schukat-Talamazzini and H. Niemann, Das ISADORA-System. Ein akustisch-phonetisches Netzwerk zur automatischen Spracherkennung. In B. Radig, Editor, *Mustererkennung 1991*, Volume 290 of *Informatik Fachberichte*, pages 251-258, Berlin, 1991. Springer-Verlag.