MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190635

Design and Evaluation of an Automatic Speech Recognition Model for Clinical Notes in Spanish in a Mobile Online Environment

Alejandro Renato, Hernan Berinsky, Mariana Daus, Miguel Fantin Dachery, Oscar Jauregui, Fernando Storani, María Laura Gambarte, Carlos Otero, Daniel Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina

Abstract

Clinical documentation in healthcare institutions is one of the daily tasks that consumes most of the time for those involved. The adoption of mobile devices in medical practice increases efficiency among healthcare professionals. We describe the design and evaluation of an automatic speech recognition system that enables the transcription of audio to text of clinical notes in a mobile environment. Our system achieved 94.1% word accuracy when evaluated on pediatrics, internal medicine and surgery services.

Keywords:

Speech Recognition Software, Mobile Applications, Electronic Health Records.

Introduction

Clinical documentation is an important task, but also time consuming, often taking up 50% of the total time of health professionals [3]. This is where the need arises for new support technologies such as speech recognition, which could allow professionals to free up time and resources. Professionals could then focus more attention on the relationship with the patient. In addition, speech recognition technology maintains the quality of clinical documentation carried out and applied according to the workflow of the professional [4].

In this work carried out in the Hospital Italiano of Buenos Aires, we describe the development and implementation of an application of automatic speech recognition (ASR) oriented to medical domain in a mobile environment.

Methods

In 1998, the Hospital Italiano de Buenos Aires (HIBA) began implementation of its Health Information System (HIS), integrating the administrative and clinical workflows with the applications in use. This system handles all clinical and administrative information, from capture to analysis. Currently, the Electronic Health Record (EHR) is a software in web format, with problem oriented architecture and is focused at the patient level.

Automatic Speech Recognition System

The ASR system was developed internally by HIBA using the open source Kaldi system [2]. For the construction of the system, both the acoustic and the language models, and the acoustic dictionary were developed as part of our work, as well as a reverse text normalization module.

The acoustic model was trained with approximately 800 hours of audio from Spanish speakers from different locations in Argentina. At the end of the training period, we obtained a Neural Networks model (NNET3 in terms of Kaldi, based on sequential models with an I-Vectors module).

The language model used in this work was constructed from the clinical notes performed at HIBA for 5 years, a total of 12 million. The notes comprise 800 million tokens, and a vocabulary close to 80,000 terms. More than 60 medical specialties were represented in the notes. For the construction of the model, the text was standardized and normalized. With this textual database, we built a trigrams model consisting of 62,000 words.

The pronunciation of the 80,000 words of clinical notes was obtained automatically by the grapheme to phoneme converter and then supervised manually. Finally, a module was built to perform the inverse normalization.

The ASR System architecture is shown in Figure 1. ASR Service supports full-duplex communication based on websockets in a scalable architecture: HAProxy distributes incoming requests on multiple independent Kaldi GStreamer Servers [1]; each server supports concurrent processing running multiple workers. GStreamer plugins (GST plugin), installed on Kaldi, executes the decoding task. This task transforms input audio signals to text transcription output based on the ASR Model. The Mobile application for Android is built with Ionic 1 framework and AngularJS (using Cordova plugins to access to the microphone).

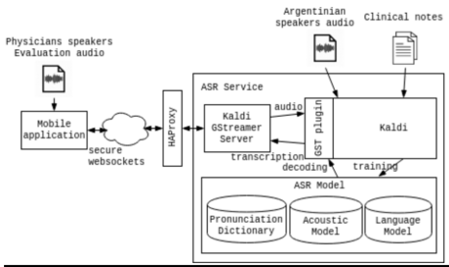

Figure 1– Automatic Speech Recognition (ASR) System architecture

Results

Over eight hours of audio recorded from the reading of clinical notes were used to assess the performance of the system.

The texts were selected from clinical notes of three specialties: internal medicine, pediatrics and surgery. The objective of the selection of texts was that several dimensions of the texts were represented: length of clinical notes measured in words, coverage of the vocabulary and less frequent phonemic combinations. To fulfill this objective, a random selection was made with the following requirements: length between 100 and 200 words, containing one of the 20,000 least frequent words, and containing at least one of the less frequent triptones found during acoustic model training. Following these criteria, 12 batches of data were obtained, each consisting at least 60 clinical notes, and were given to read to HIBA physicians who knew the vocabulary of each specialty.

The speakers were selected from the set of resident physicians at HIBA. Of the twelve residents, eight were women and four were men. One of the speakers was not a native Spanish speaker (Brazilian Portuguese), while another is a native Spanish speaker, but with a different dialect (Colombia). From the remaining speakers, nine speak Rio de la Plata Spanish as their native dialect.

The texts were read in a silent environment, using the mobile client designed for the final application. The logs of the system were collected, along with the original transcription and the audios of the system.

The results of the system without tuning were 91.7% Accuracy, 8.3% WER (Word Error Rate), and 93.4% Percent Correct.

Table 1 shows the results obtained after tuning the system. WER is determined by subtracting Accuracy from 100. Accuracy is different from Percent Correct Word, since it includes both inserts and deletions. The second column in Table 1 gives the percent values while the third column shows absolute values.

Among the serious errors, which are the most difficult to control, are the numerical quantities. Of the 5666 numerical expressions that appear throughout the recorded corpus, there were 307 errors in recognizing numbers (5.41%). Although it seems that the error rate is not high, the occurrence of one of these errors in some contexts can threaten the patient's safety.

Finally, a comparison was made with a set of typed clinical notes of similar characteristics extracted randomly from the set of notes. An analysis of the errors found was then made. The analysis shows that the degree of correct words is 89.2%, versus 95.8% obtained by the ASR system.

Table 1– Results Evaluation

Measurements	Percent	N Words
Word Error Rate	5.9%	2745
Percent Correct	95.8%	44406
Substitution	2.9%	1354
Deletions	1.3%	607
Insertions	1.7%	784
Word Accuracy	94.1%	

Conclusions

We conclude that our mobile ASR system for taking clinical notes has been shown to perform acceptably. Other studies are underway to verify parameters such as: tool safety, clinical applicability, performance compared to the traditional method and quality of the record.

References

[1] T. Alumäe, Full-duplex Speech-to-text System for Estonian, in: *Proceedings of the Sixth International Conference Baltic HLT 2014,* G.G. A. Utka, J. Kapočiūtė-Dzikienė, ed., IOS Press, 2014, pp. 3-10.

[2] A.G. Daniel Povey, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, Karel Vesely, The Kaldi speech recognition toolkit, *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding* (2011).

[3] N.H. Solutions, Accuracy & completeness of clinical documentation, in, 2015.

[4] W.D.A. T.H. Payne, J.A. Markiel, K. Lybarger, and A.A. White, Using voice to create hospital progress notes: Description of a mobile application and supporting system integrated with a commercial electronic health record, *Journal of Biomedical Informatics* **77** (2018), 91-96.

Address for correspondence

Alejandro Renato
alejandro.renato@hospitalitaliano.org.ar