MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Development of Medical Conversation ASR system in Argentine Spanish

Sonia Elizabeth Benitez , MD, MSc

Head of Research & Technology Innovation, Department of Health Informatics

Hospital Italiano de Buenos Aires

Twitter
@soniaebenitezok

AIDH IMIA

Presentation Journey

1. Introduction
2. Methods
3. Results
4. Discussion & Conclusions

Introduction

Background

Speech Transcription Physician/ Patient

-Medical Reports have 71-73% of accuracy when compared to recorded audio
-Commercial engines: 35 to 65% of WER*
-Google has a systems with 18% WER trained in medical reports and physician/patients interviews, only in English.
-Lack of System in Spanish

*Word Error Rate

Objectives

To evaluate the performance of the physician-patient conversation speech to text transcription system developed at HIBA

To compared with a NeMo Conformer system for Spanish (E2E) and the Google ASR system in Spanish

Introduction

Difficulties

- Spontaneous speech: false starts, filled pauses, and unfinished sentences
- Speaker distance from the microphone or the low computer recording quality, voice overlapping, and ambient noise.
- Problems with diarization -which segments the recording into speech turns per speaker-
- Lack of medical vocabulary

Hospital Italiano de Buenos Aires

- Academic hospital
- EHR & HIS **in-house developed** for more than 20 years
- Certified & accredited by several organizations

HIMSS Analytics STAGE 7

CAP ACCREDITED COLLEGE of AMERICAN PATHOLOGISTS

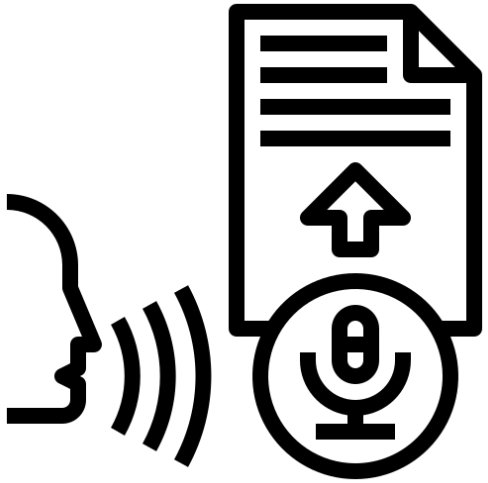
Hospital Acreditado por Joint Commission International

Methods

- To develop the physician-patient conversation speech to Text Transcription System.
- To evaluate the performance of the system, with a set of 208 teleconsultations recorded during year 2021, representative of hospital environment and patient speech.
- The system proposed was compared with a NeMo Conformer system for Spanish (E2E) and the Google ASR system in Spanish.
- The evaluation was measured in WER for the ASR system, in Recall and F1 for recognized medical entities

Methods

- The overall system performance depends on the automatic recognition system as a critical element
  - acoustic training
  - the language model building



Methods

- Acoustics Training
  - A large corpus (42,000 hours) representative of the phonetic variation and dialects of Argentina was collected
  - 6000 hours of speech representative of spanish argentine dialects, with preference in spontaneous speech.
  - A factored TDNN* model was trained using Kaldi Toolkit.

*Time Delay Neural Network

Methods

- Language Model Training
  - A corpus of medical reports made at the HIBA between 2017 and 2021 (anonymized)
  - 144 million sentences and a vocabulary of 91702 words was selected (frequency of use greater than 20 times in 5 years) was used

Methods

- Language Model Training
  - Pronunciations of recognized spanish words were automatically generated by training the Phonetisaurus system
  - Foreign words, proper names, and abbreviations were corrected by hand.

Methods

- Language Model Training
  - The search criteria were the occurrence of first and second person pronouns and verbs, which reflected speech in a verbal interaction
    - 20 million sentences and 177 million words
  - The dictionary in total has 388649 words
  - A quadrigram model was built with the corpus

Results

- Test set: 208 teleconsultations. 16 hours & 35 minutes of audio
- Participants: 12 physicians & 202 patients

Table 1. ASR WER and Accuracy, and Entity Recognition Precision Recall and F1

Model	Kaldi TDNN	NeMo Conformer Spanish	Google API Spanish
Accuracy/WER	72.5 % / 27.5	64 % / 36	39 % / 61
Precision (Entities)	0.98	0.97	0.86
Recall (Entities)	0.87	0.78	0.42
F1 (Entities)	0.80	0.42	0.28

Results

Kaldi TDNN, system

-errors are concentrated in connectors (articles, prepositions) or in morphological variants (plural vs singular)

-lexemes are preserved

NeMo Conformer

- errors creating non-existent words that have pronunciations similar:
“Amlodipine”=>lodipine” or “anglodipine”
“diabetic” => “llabetico”
“colesterol”=> coletrol”

Google ASR Spanish

-It recognized medical words as existing names, such as “Parkinson's”=>parking
”obstructions”=> “demosthenes”

Discussion & Conclusion

- The proposed system has conditions for improvement
  - To record on different channels in order to avoid voices overla
  - To improve audio quality using appropriate microphones
  - To perform fine tuning with the audios
  - To adapt the language models to the domain
  - To use data augmentation strategies or synthetic data creation

Development of Medical Conversation ASR system in Argentine Spanish

Thank you !
Any questions?

alejandro.renato@hiba.org.ar
investigacion.informatica@hiba.org.ar

HOSPITAL ITALIANO
de Buenos Aires
Departamento de Informática en Salud