ELSA SPEAK

Language resources management at ELSA

Xavier Anguera, Cofounder & CTO
# Table of Content

1. **ELSA Introduction**
2. **Elsa Speak**
3. **Data projects**
1 ELSA Introduction
About me

- 8 years in a big telco as multimedia scientist
- Moved to Lisbon in 2015, started a solo-startup around reading e-books
- Joined ELSA as co-founder, Chief scientist and CTO end of 2015

Xavier Anguera
xavier@elsaspeak.com
a personal mobile coach that improves users' English pronunciation and speaking skills so they can speak clearly, fluently, and confidently like a native English speaker

state-of-the-art speech recognition technology to pinpoint errors, then gives accurate feedback to our users on how to improve.
What we do

• We develop **A.I. and speech technology** for spoken pronunciation assessment

• We detect pronunciation errors at phoneme level and **give detailed feedback** and phonetic hints on how to overcome them.
What we are

• Our products:
  • IOS and Android applications: More than 4M downloads
  • Language assessment API available
  • Online teacher dashboard
• Our technology:
  • Speech recognition technology developed in-house
    (Speech research team in Lisbon)
  • Robust and scalable backend tech stack
• Our team:
  • 20 people split across 3 continents (US, Portugal, India and Vietnam)
  • Multicultural backgrounds in a Silicon Valley startup culture
ELSA for students

ELSA for schools

ELSA for Taxi drivers

ELSA for hotel staff
San Francisco (HQ)
- Product
- Design

Portugal
- Engineering
- Research

India
- Engineering

Vietnam
- Engineering
- Customer support
- Biz dev.

Early 2015: ELSA started by Vu Van

Late 2015: I join as CSO & CTO

March 2016: ELSA wins SXSW education Startups competition

March 2016: IOS and Android versions

Jan 2017: ELSA 2.0 release

Jan 2018: 1M downloads in Android

March 2018: 3.2M funding by Monks Hill

Feb. 2019: 7M A round funding led by Gradient Ventures
2 ELSA Speak
Our objective is to help our users improve their communication skills in American English.

- Current exercise types:
  - Pronunciation
  - Listening
  - Conversation
  - Intonation (word stress)
ELSA Speak - pronunciation exercise

Let's start with the basics

- Lesson 1 - Listening teaser
  - Easy
- Lesson 2 - Short /s/ vs. long /i/ - 101
  - Easy
- Lesson 3 - Sentence pairs
  - Easy

Reading

- Lesson 4 - Reading 101
  - Easy
- Lesson 5 - Reading Difficulty
  - Easy
- Lesson 6 - Book genre
  - Medium
Oscar Nominee

Excellent!

Smile, as you say the word 'nominee' with a long vowel 'ee' at the end.

International Phonetic Alphabet (IPA)

/ˈskərˌnomiːni/
● 16 English pronunciation skills
● 35 content modules
● 1001 Lessons
● 5280 exercises
  ○ Pronunciation: 4149 exercises
  ○ Conversation: 458 exercises
  ○ Listening: 213 exercises
  ○ Word stress 460
Dictionary data

- Free text, entered by users or selected from a recommended list
- 230K different sentences since we started tracking them (approx. 6 months)
- Not all sentences are correct (politically and grammatically)
Data collection

Speech processing server

Amazon S3
Data collection (raw data)

- ~ 1 Million exercises/day
- ~ 25 TB of audio+metadata stored so far
Data collection (cleaned up data)

- About 7TB of “clean” data so far

Cleaning steps:

- Is all text correctly aligned?
- Is overall alignment score good enough?
- Is SNR good?
- Are there no more than 1 consecutive mispronounced phonemes?

Then subsets of the above are selected for the different tasks
Audio samples: assessment test sentences

1. This summer I will visit a new country with two of my best friends.
2. We will go sightseeing and stay at a resort by the ocean.
3. It will be very good weather; we look forward to swimming and sunbathing.
4. ...
3 Data projects
Our first acoustic model was US English data
  ○ Out of domain
  ○ non-matching acoustic conditions

We feed the most accurate sentences into the training data
  ○ Same model topology, using out-of-domain and in-domain data

CAUTION: This is not ASR, we want to keep error detection precision/recall
  ○ Only the top audios are used

PROJECT 1: Acoustic models adaptation
ASR systems perform badly when non-native accents are using them. Availability of non-native data can improve ASR recognition. Proof of concept: training an ASR LVCSR model with Vietnamese data:
- 500h of training data in Vietnamese English, from ELSA + 500h from Librispeech
- 4h of test data of Vietnamese English, from ELSA. Different sentences and speakers as training.

Librispeech
500h

GMM/HMM training

DNN/HMM training

decode

Non-native English
500h

Non-native English
4h
● Results obtained on 8KHz data, further tests on 16KHz data showed better results over all.
● Very impressive results from API-1 whose acoustic models are not adapted to Vietnamese
● Other possible adaptation approaches we plan to test:
  ○ Chain models with mixed training data
  ○ Transfer learning with Librispeech + non-native
  ○ Multitask learning
We can extract useful information from pronunciation errors detected by our system.

We compare the expected pronunciation with the pronunciation recognized by a phoneme decoder.

Experiment in Vietnamese:
- Processed ~300k user audios
- Extracted ~190 most common phoneme confusions + the information about the contexts in which they occur.
- Hypothesis: These should match the literature
- Must: Need to manually check whether the confusion is due to an error from the user or an error from the classifier.
“HI ELSA, nice to meet you!”

“Judgmental”
● ELSA is helping many users improve their pronunciation skills in English
● We collecting humongous amounts of non-native data
  ○ But it is non-labelled, and sometimes dubious in quality
● We are using this data extensively to improve our detection of pronunciation errors in our app
  ○ Lots of works and techniques still to try
● We are hiring!
THANK YOU

Xavier Anguera
xavier@elsaspeak.com