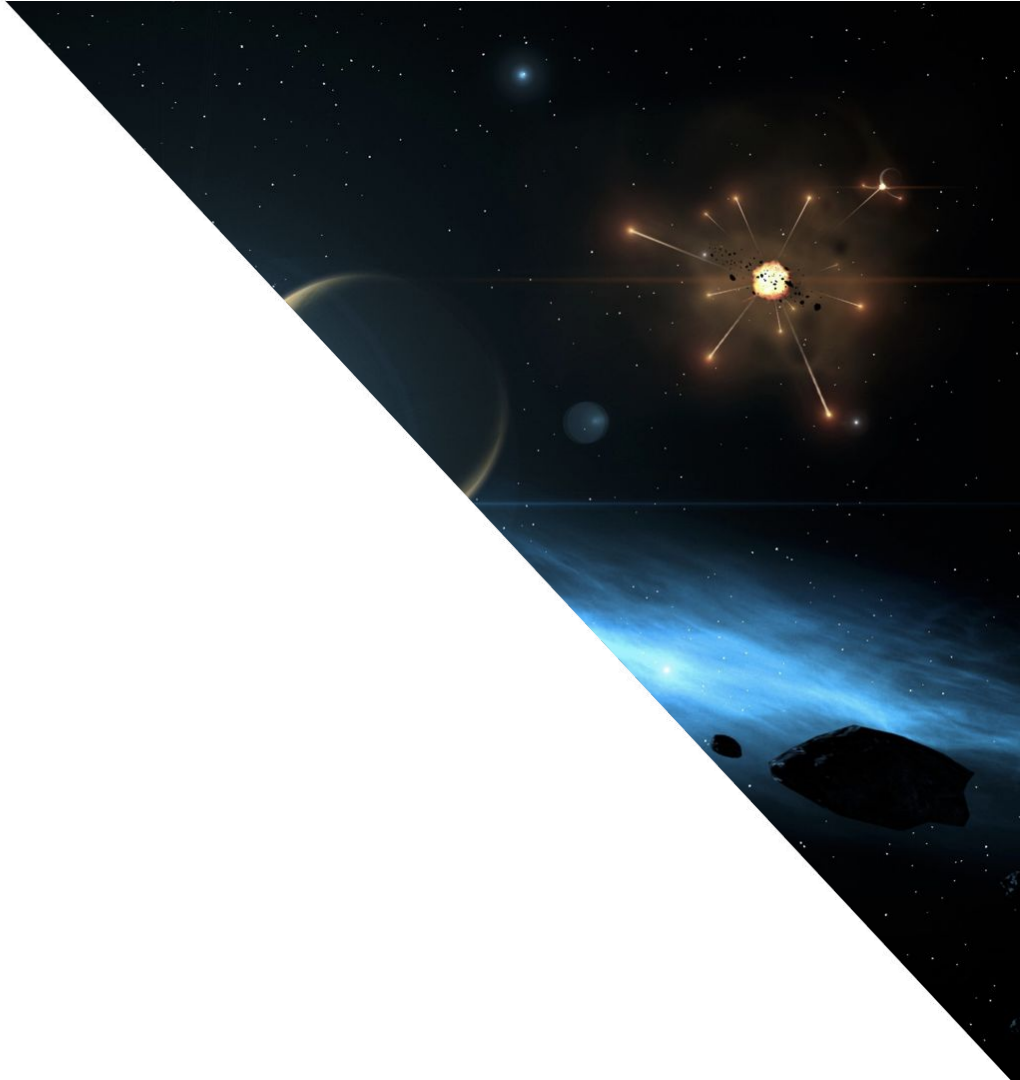
ELSA SPEAK

Language resources management at ELSA

Xavier Anguera, Cofounder & CTO

Table of Content



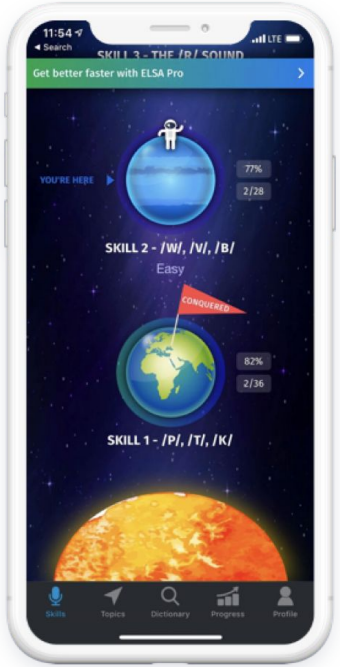
1 ELSA Introduction

About me

- Ph.D. (2006) in Speech processing from UPC, Barcelona
- 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

Who we are?

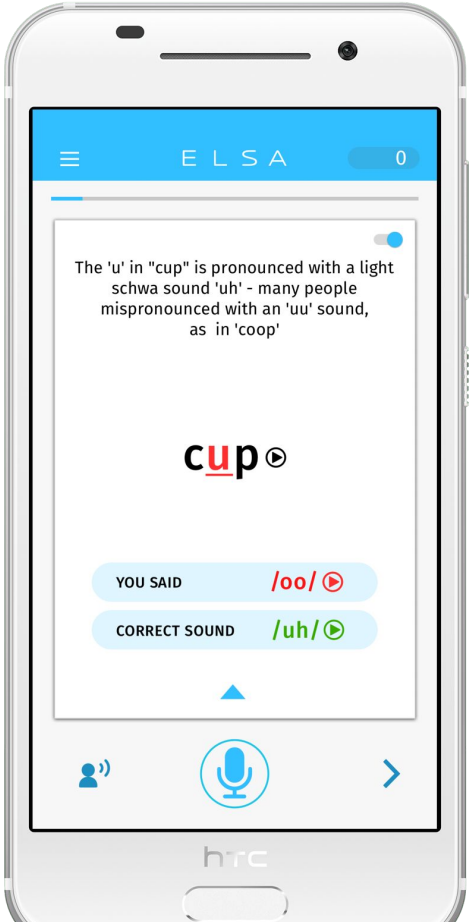


ELSA SPEAK

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.

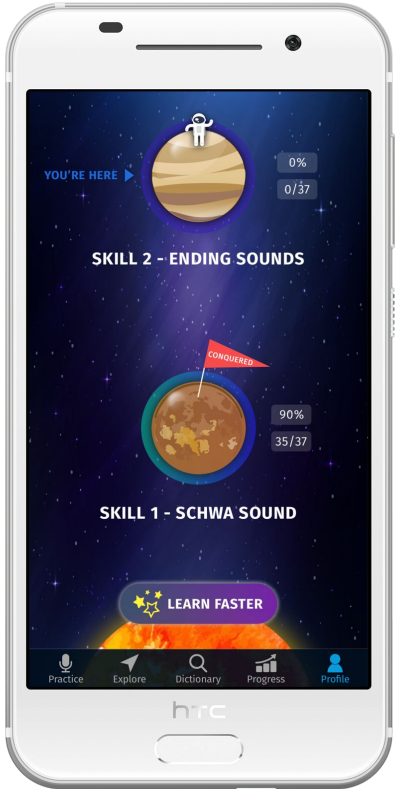
What we do



- 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

Available on the iPhone App Store

ANDROID APP ON Google play

Products

ELSA for students

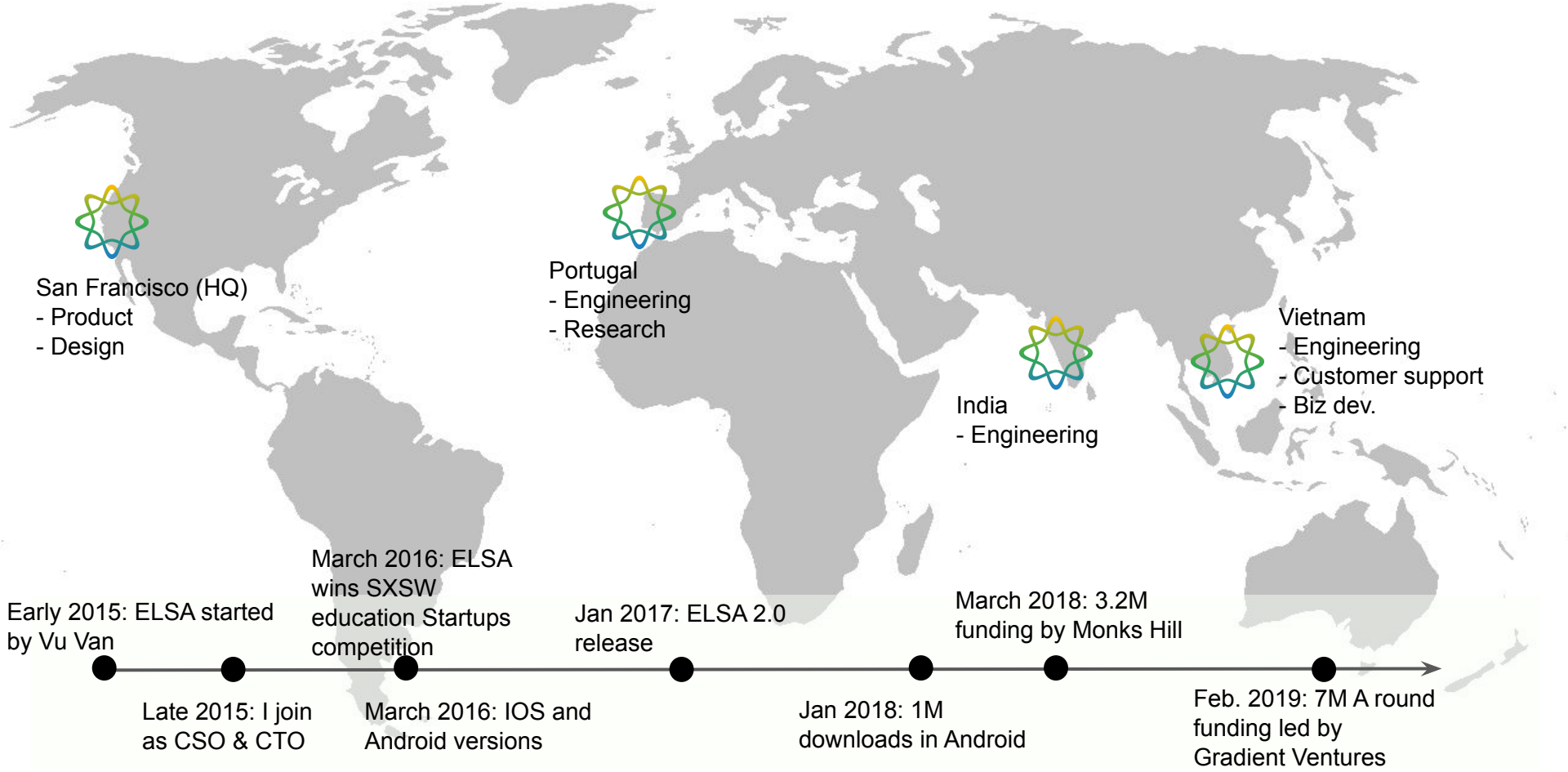
ELSA for schools

ELSA for Taxi drivers

ELSA for hotel staff

ELSA

Locations

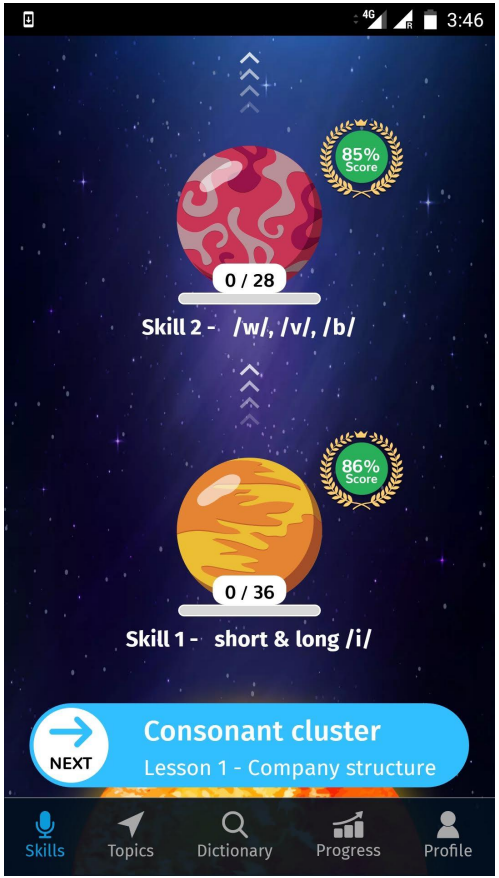


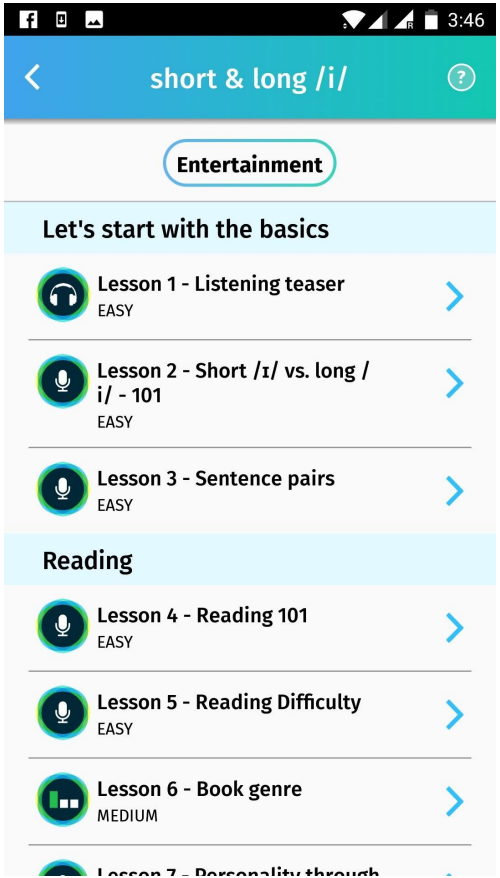
2 ELSA Speak

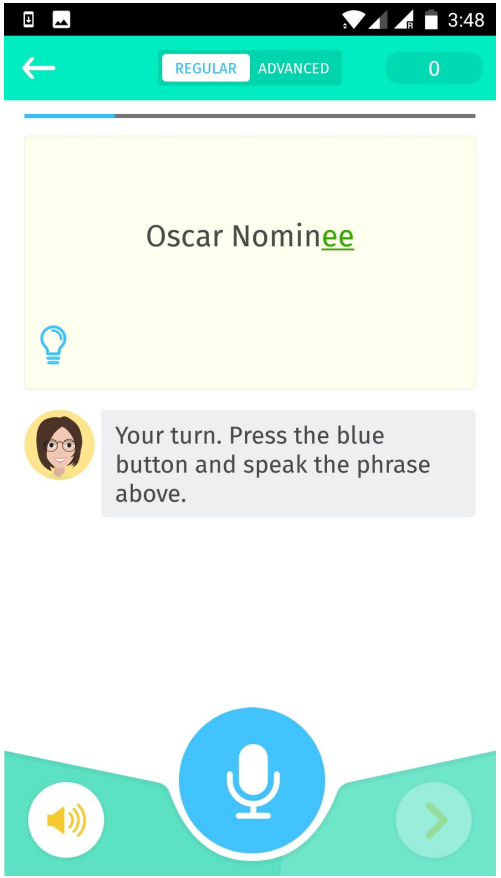
Our objective is to help our users improve their communication skills in American English.

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

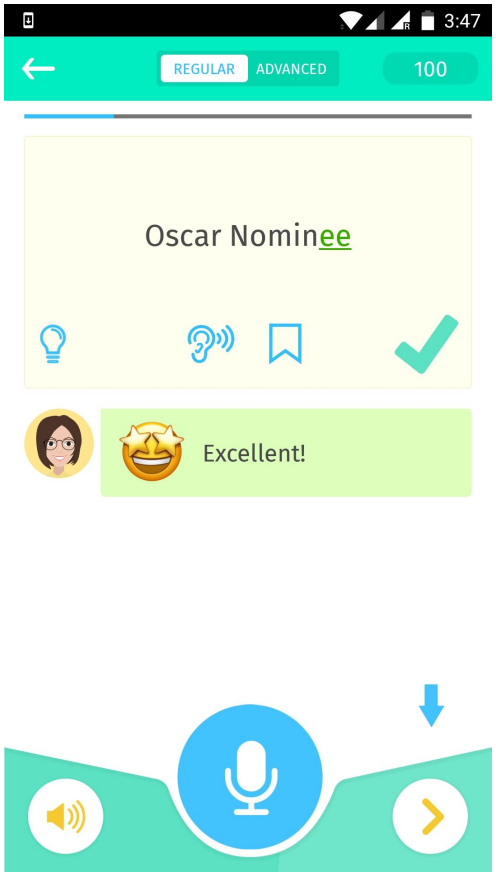
ELSA Speak - pronunciation exercise

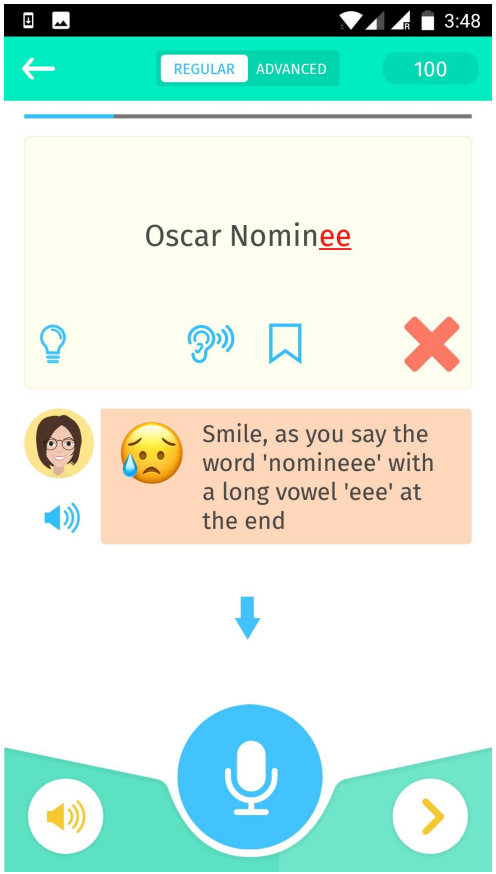


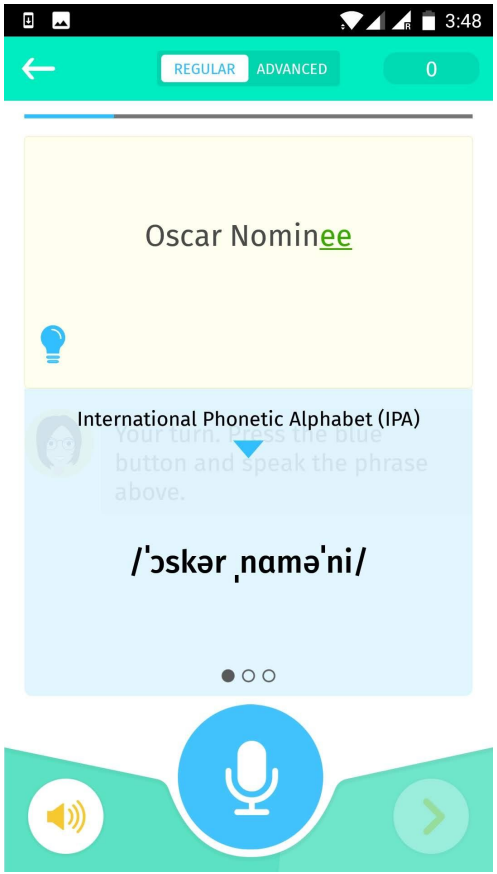




ELSA Speak - pronunciation exercise

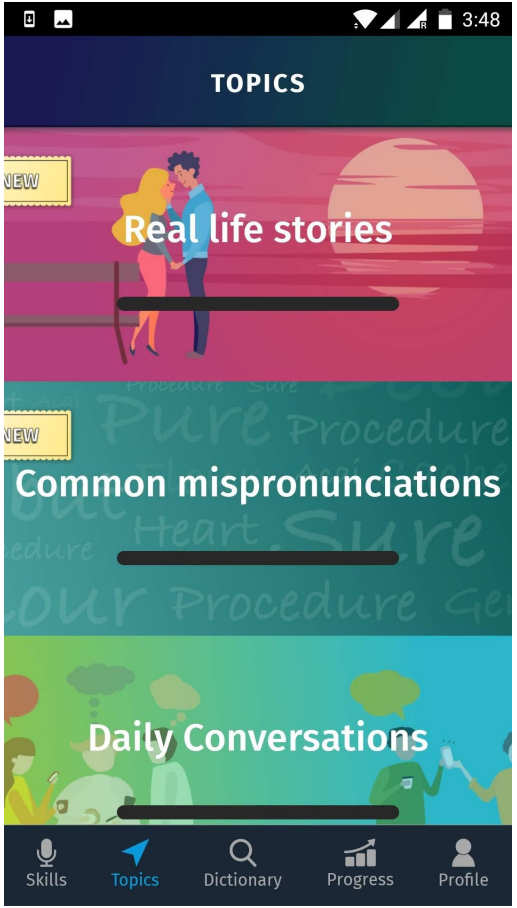






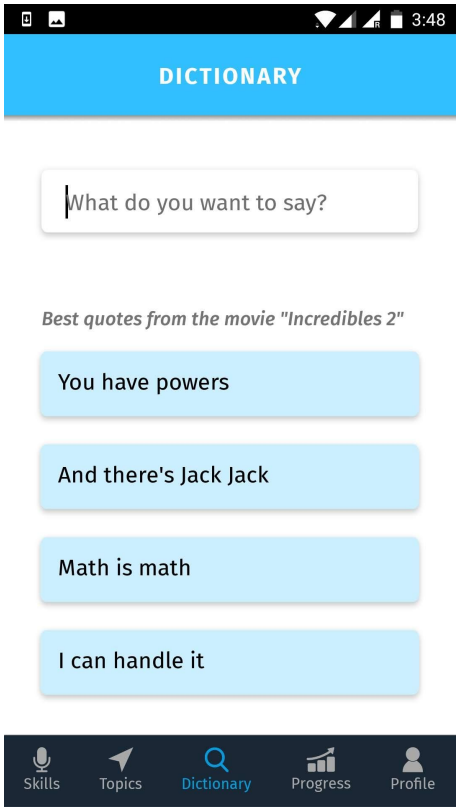
ELSA Speak curricula (March 2019)

- 16 English pronunciation skills
- 35 content modules
- 1001 Lessons
- 5280 exercises
  - Pronunciation: 4149 exercises
  - Conversation: 458 exercises
  - Listening: 213 exercises
  - Word stress 460

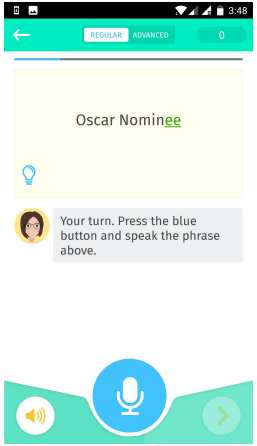


Dictionary data

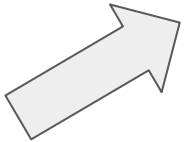
- Free text, entered by users or selected from a recommended list
- 230K different sentences since we started tracking them (approx. 6 months)
- Not all sentences are correct (politically and grammatically)



Data collection

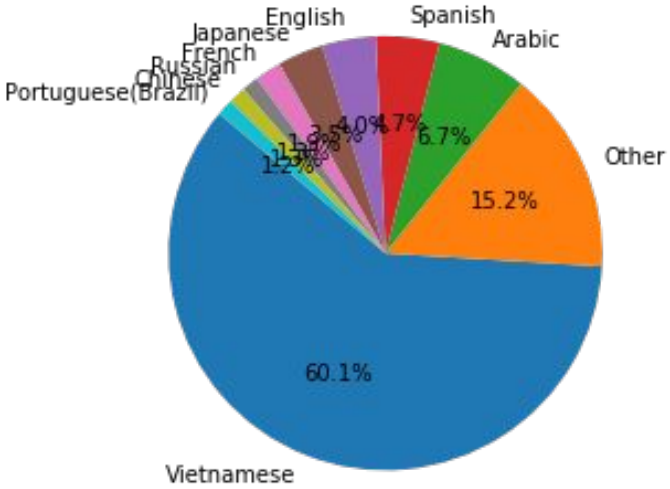


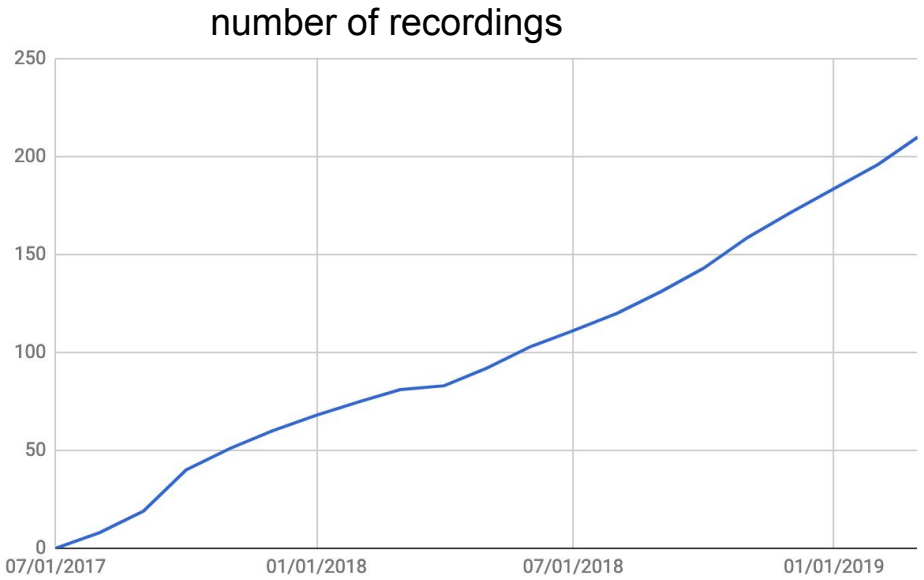
Speech processing server



Data collection (raw data)

- ~ 1Million exercises/day
- ~ 25TB of audio+metadata stored so far





Data collection (cleaned up data)

- About 7TB of “clean” data so far

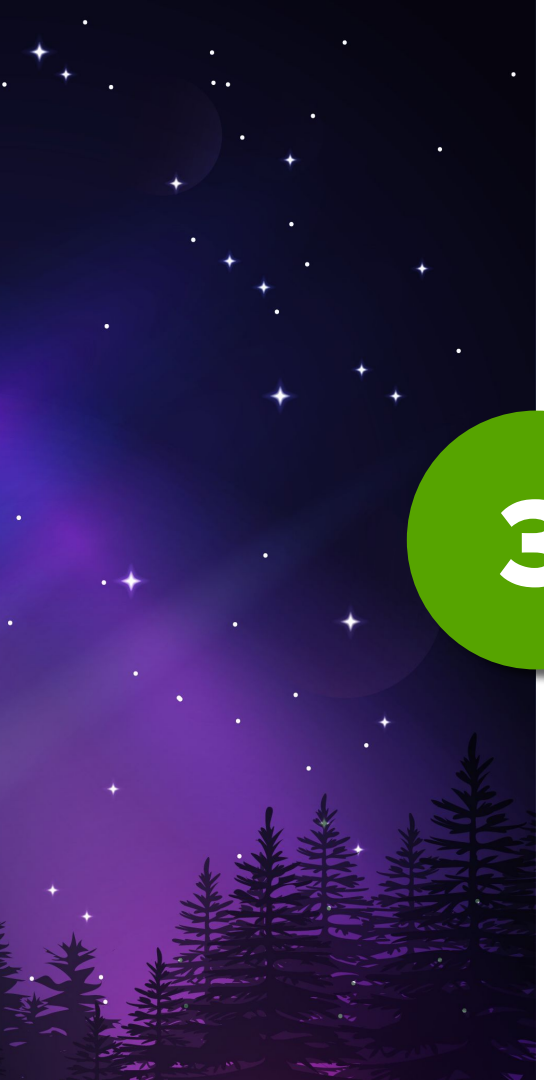
Cleaning steps:

- 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

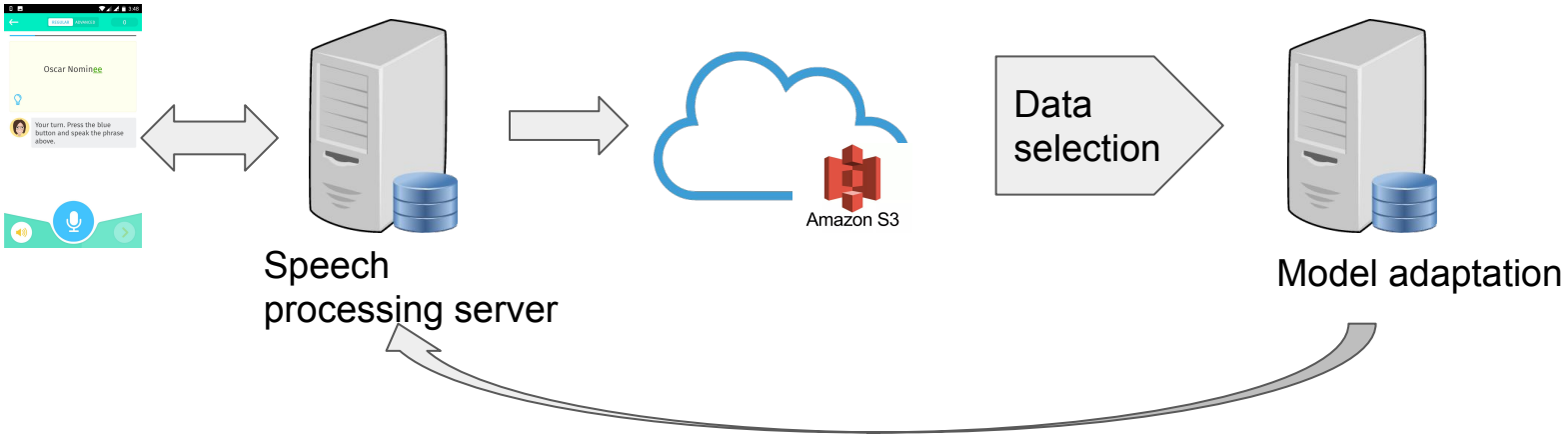
1. This summer I will visit a new country with two of my best friends.
2. We will go sightseeing and stay at a resort by the ocean.
3. It will be very good weather; we look forward to swimming and sunbathing.
4. ...



3 Data projects

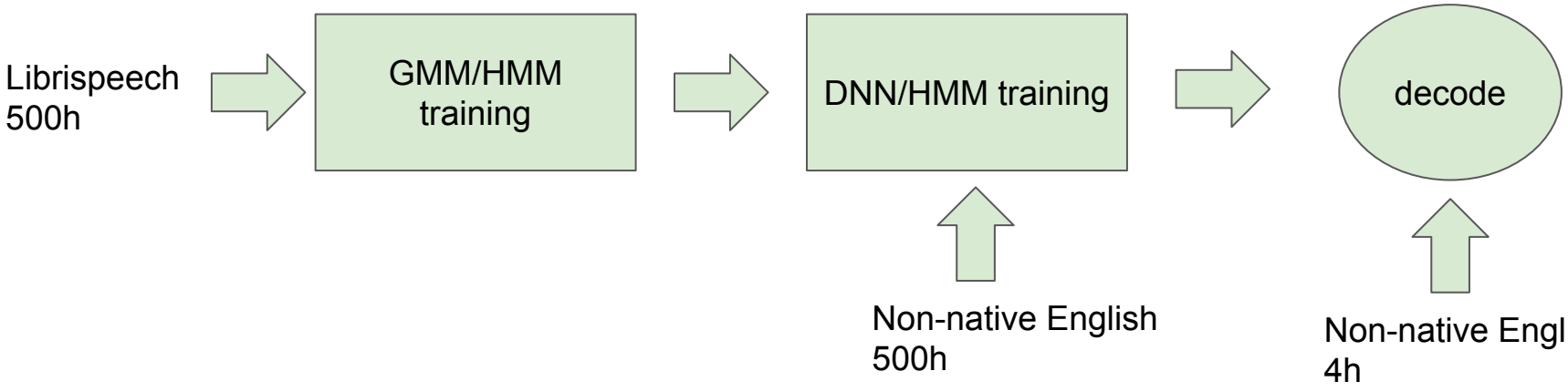
PROJECT 1: Acoustic models adaptation

- 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 2: Non-native English ASR

- 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



PROJECT 2: Non-native English ASR

System	WER
ELSA	25.26%
3rd party API-1	26.05%
3rd party API-2	44.29

- 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

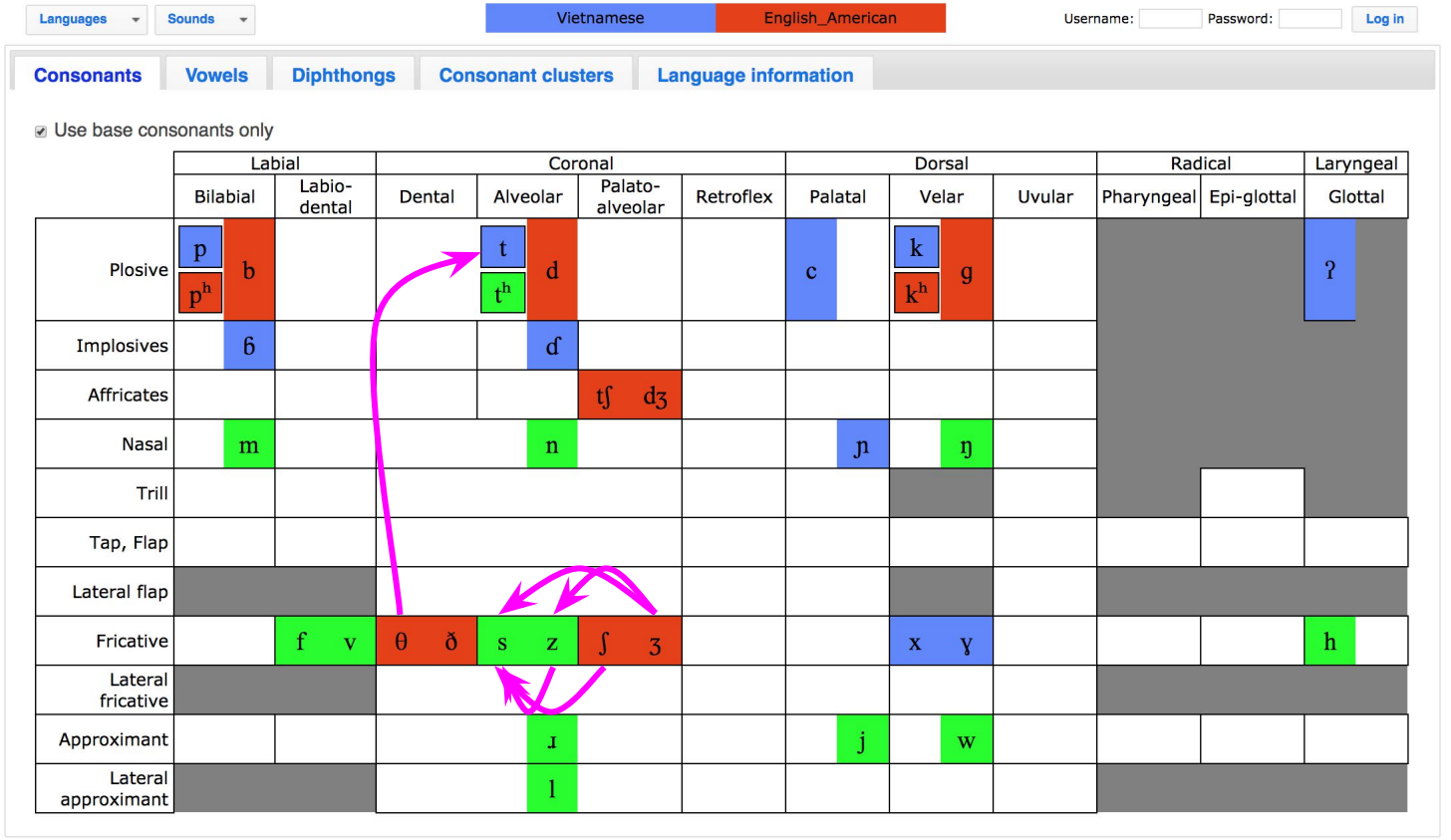
ELSA

PROJECT 3: Analysis of common mispronunciations

- 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 of an error from the classifier.

Vietnamese

Expected	Realized	Frequency
ZH	Z	0.2196
SH	S	0.192
Z_E	<eps>	0.185
D_E	<eps>	0.1722
TH	T	0.152
Z	S	0.1403
T_E	<eps>	0.1396
IH	IY	0.1395
ZH	S	0.1372
R	<eps>	0.1361
Y_I	IY	0.1335
ZH_I	SH	0.1329
G	<eps>	0.1257
NG_I	N	0.1256
AO_I	AA	0.1205
DH_E	TH	0.1189
K	<eps>	0.1175
DH_E	TH	0.1189
K	<eps>	0.1175

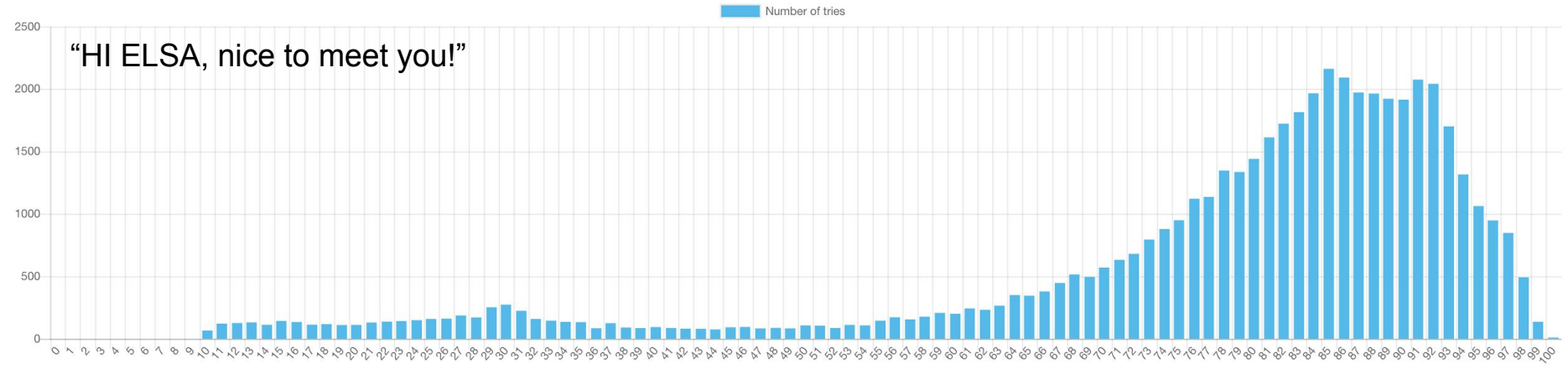


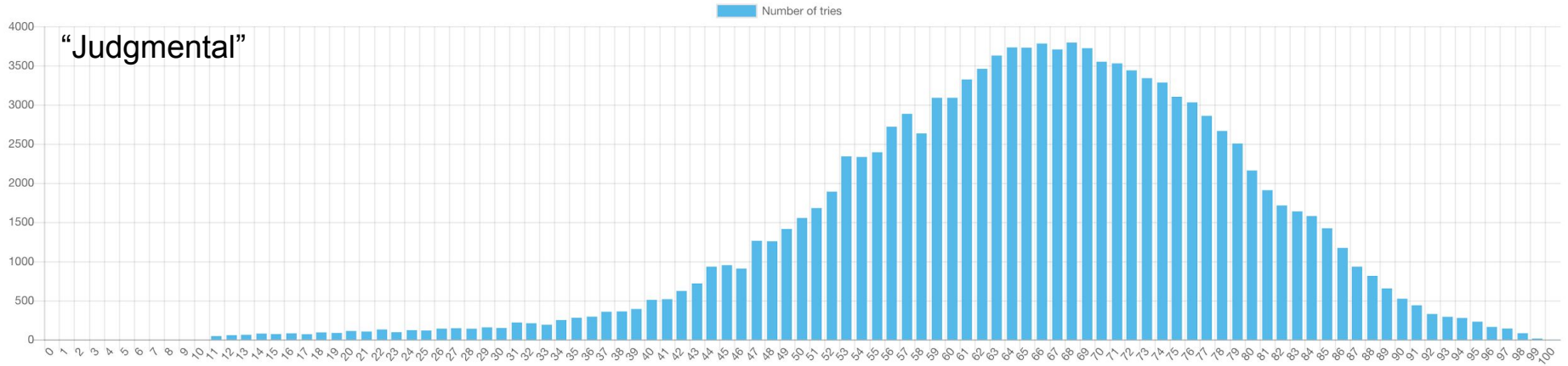
NTNU – Trondheim
Norwegian University of Science and Technology

l1-l2map

https://l1-l2map.hf.ntnu.no/l1-l2map

PROJECT 4: exercise difficulty





Summary

- 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