Table of Content

Empathy

Backend Development

Frontend Development

Table of Content

Empathy

Finding the Problem

Task 1

Decided to focus on communication problem

Task 2

Narrowed down to cerebral palsy patients

But we had a hard time finding a solution that allows easy testing of user-interaction without eIRB approval

Task 2 Reapproached

Shifted to the refugee population

Identified their language learning problem that is common even in our class

Empathy, ideation, brainstorm

Performed research on the population

Had a better grasp of their major challenges and crystalized into ideas

Key Discoveries

Inconsistent feedback from underqualified language teachers

English as second language (ESL) teachers are usually non-native speakers and require less years of training than language arts teachers.

Verbal skills are usually the slowest in development

Due to the shame that children feel when they speak to others but cannot be understood. Older children identifies their different accents and feel self-conscious. Then, they tend to speak to only those who speak their original language.

Refugees at different ages learn for different purposes

Table 1. How and why are the participants learning English in Lebanon.

Participant	Gender	Age	Occupation	Learning Method	Motivation	Internet Access Method	Familiarity with Mobile Technology
P1	Female	16	N/A	English sessions	Study in Canada	Father's phone	Unfamilar
P2	Female	14	Student	At school	Succeed at school	Father's phone	Unfamilar
P3	Male	15	Student	At school	Succeed at school	Own phone	Familar
P4	Male	23	Plumber	Duolingo	Move to Canada	Own phone	Familar
P5	Male	15	Mechanic	English sessions	Leave Lebanon	Own phone	Familar
P6	Male	18	Mechanic	English sessions	Leave Lebanon	Own phone	Familar
P7	Female	25	N/A	Dictionnary	Help her children	Husband's phone	Unfamilar
P8	Male	18	Student	At school	Succeed at school	Own phone	Unfamilar

Table 2. How and why are the participants learning German in Germany.

Participant	Gender	Age	Occupation	Learning Method	Motivation	Internet Access Method	Familiarity with Mobile Technology
P1	Female	54	N/A	German school	Find employment	Own phone	Unfamiliar
P2	Female	29	N/A	German school	Enroll at university	Own phone	Familiar
P3	Female	26	N/A	German school	Enroll at university	Own phone	Familiar
P4	Male	26	Student	Applications	Stay in Germany	Own phone	Familiar
P5	Male	31	Student	Applications	Stay in Germany	Own phone	Familiar
P6	Male	30	Student	Applications	Get naturalized	Own phone	Familiar
P7	Male	39	N/A	German school	Find employment	Own phone	Familiar
	Male	35	N/A	German school	Find employment	Own phone	Familiar
	Male	29	N/A	German school	Find employment	Own phone	Familiar
	Male	39	N/A	German school	Find employment	Own phone	Familiar

"Language Learning Tool for Refugees: Identifying the Language Learning Needs of Syrian Refugees Through Participatory Design"

Table of Content

Empathy

Backend Development

Frontend Development

Backend Models

- **Acoustic Model**
  - Recognize the phoneme
  - confidence level of pronunciation
- **Decoding Model**
  - force-alignment for the speech
  - based on the phoneme posteriorgram
  - and the reference phoneme sequence
- **Scoring Model**
  - Apply the Goodness of Pronunciation (GOP)
  - Get the score of the users

Backend Development

Acoustic Model

Trained for phoneme recognition and their confidence level on pronunciation

Purpose 🚀

Model structure: state-of-the-art model (subsampling Factorized Time Delay Neural Network with Lattice-Free Maximum Mutual Information criterion - F-Chain-TDNN)
Phonetic transition: Hidden Markov Model

Model ✏️

Data

Librispeech, a popular open-source corpus with 1000 hours of speech

Mel-frequency Cepstrum Coefficients (MFCC) are extracted for input features. Then, we gradually feed the data and add feature transformation techniques to train a **base GMM-HMM model**

Features

Features

F-TDNN model with 17 Factorized TDNN layers (each with linear + central + scaling layers) besides an LDA affine layer next to the input and a linear layer next to the output.

Backend Development

Decoding Model

Force-Alignment

based on the phoneme posteriorgram and the reference phoneme sequence

Dynamic Time Warping

without using linear alignment since it cannot solve the optional silence between phonemes

Decoding Graph

regarding the optional silences between phonemes and use the Viterbi algorithm to find the best path

Backend Development

Goodness of Pronunciation (GOP)

$$GOP(p) = \frac{log(P(p|O))}{L(O)} = \frac{log(\frac{P(O|p)P(p)}{\sum_{q \in Q} P(O|q)P(q)})}{L(O)}$$

where p is specific phoneme, O is the related audio sample to p, and q is all other phonemes.

NOTE: GOP is always interpreted as binary classes to determine whether the phoneme is accepted or rejected with phone-dependent thresholds employed

Backend Development

Performance Evaluation

The Pronunciation Scoring Experiments

Method-Rater	Pearson-1	Pearson-2	Spearman-1	Spearman-2	MIC-1	MIC-2
Rater 1	\	0.573	\	0.573	\	0.276
GOP	0.425	**0.297**	0.370	**0.315**	0.182	0.143

We average the pronunciation score on phonemes for each sentence and compute the relevance score (i.e. Pearson coefficients, Spearman coefficients, and mutual information coefficients). The results showed that **our system has been similar to human raters**.
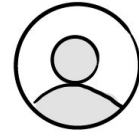
Table of Content

Empathy

Backend Development

Frontend Development

Wireframes

Fun-EMES

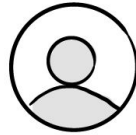The English learning of the future

Repeat after me...

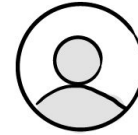This is a sample sentence.

Next

Wireframes

Grading your performance...

Here is how you did...

Listen Again...
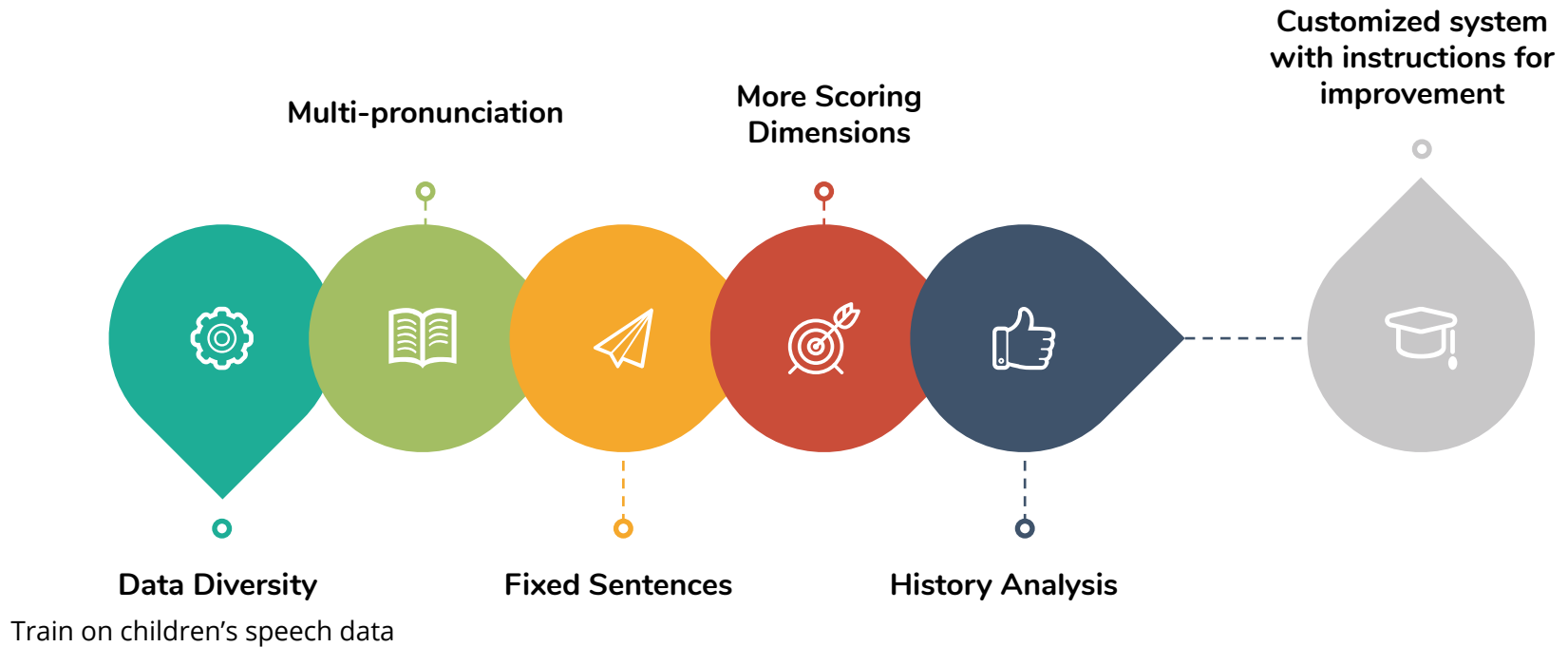
This is a sample sentence.

78

Awesome!

Repeat

Next

Frontend Demo

Future Directions

Multi-pronunciation

More Scoring
Dimensions

Customized system
with instructions for
improvement

Data Diversity

Train on children's speech data

Fixed Sentences

History Analysis

Thank You

Any Questions ?