Vowel Tuner

Soklong HIM Nora LINDVALL Maxime MÉLOUX
Jorge VASQUEZ-MERCADO

NLP M2

Software Project
Feb. 7, 2023

UNIVERSITÉ DE LORRAINE

Outline

Introduction

Our idea

Aim

Help language learners improve their pronunciation of French vowels

Idea

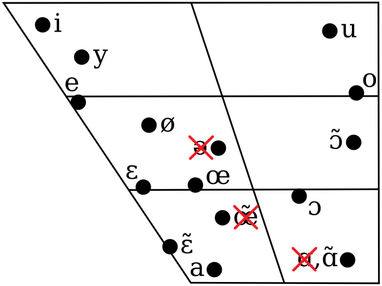
Make a web application that can identify a vowel spoken by the user, and provide personalized feedback.

Background

- *Speaking* is one of the tougher skills to master in terms of language learning [1]
- Many learners superimpose the phonetic inventory of their L1 onto L2 (phonetic substitution) [2]
- Comprehensibility can be increased if learners are made aware of phonetic boundaries in their target language.

Vowel phonemes

- /ɑ/ →/a/
  "pâte" = "patte"
- /ə/ →/ø/or/œ/
  "sur ce" /syʁsø/
  "prenait" /pʁœnɛ/[3]
- /œ̃/ →/ɛ̃/
  "brun" = "brin"[4]

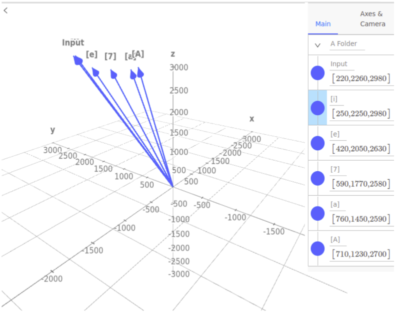


Methodology

Reference approach

Input:

- Formants
  - F1 - openness
  - F2 - frontness
  - F3 - lip rounding
  - F4 - nasality/paranasality
- Reference formants



Reference approach

Input:

- Formants
  - F1 - openness
  - F2 - frontness
  - F3 - lip rounding
  - F4 - nasality/paranasality
- Reference formants

Prediction:

- Closest vowel in formant space
- Per-formant weight
- Improved using standard deviation

Bad results (<38% accuracy), difficult to finding good reference formants

→ Quickly abandoned

Linguistic approach

Input:

- Formants F1-F4
- Phonetic context
- Speaker gender

Classifiers:

- Decision Trees
- K neighbors
- Multinomial Logistic Regression
- Random Forests
- Multilayer Perceptron
- Extra Trees
- Bagging
- Stacking

Neural Approach

- From audio to mel-spectrogram
- Images normalized and re-scaled
- Fed into CNN
- Softmax - probability of each vowel
- Vowel with highest probability is chosen

Vowel Extraction

Problem: The user says an entire word

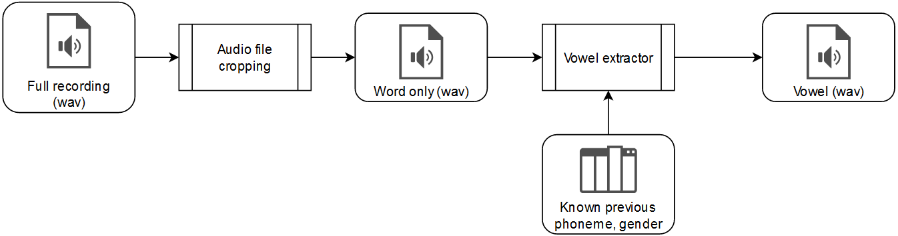


Figure: Vowel Extraction pipeline

Datasets

- An informal corpus, recorded in real-life conditions, 8 students (6 male and 2 female), of which 5 native French speakers and 3 non-native learners.
- A subset of the InterFra corpus[1], 2 non-native speakers, 2 native speakers, 225 vowels.
- The All Vowels corpus, native French speakers (34 female and 33 male speakers), 5,755 vowels.
- A testing corpus, French speakers in testing conditions, 900 vowels (see later)

[1] https://spraakbanken.gu.se/en/resources/interfra

Application

Application

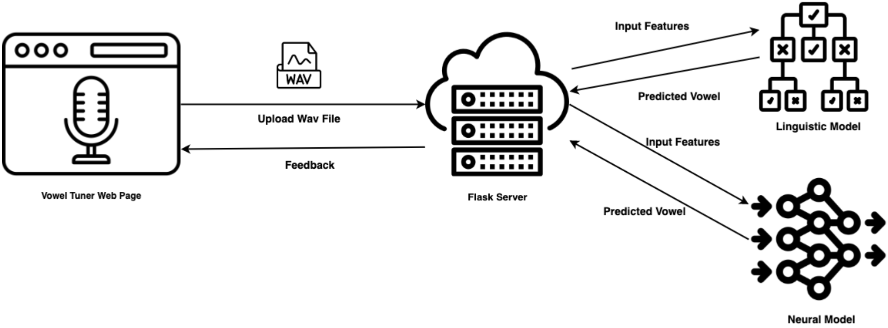


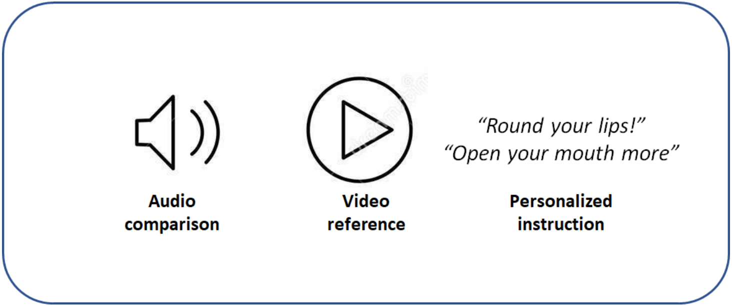
Figure: Vowel Tuner Application Diagram

Demo

Demo time!

Feedback

Feedback in three forms:



Personalized feedback

Each vowel was tagged with four attributes:

- Openness
- Frontness
- Lip-rounding
- Nasality

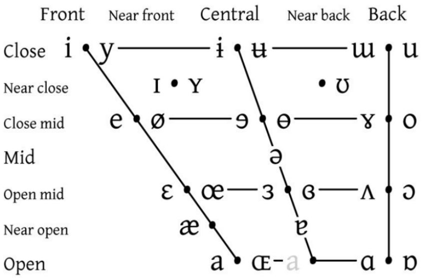


Figure: IPA vowel chart, courtesy of the International Phonetic Association.

Results and Discussion

Linguistic classifier

Classifier	Accuracy
Decision trees	74.25%
K neighbors	79.40%
Logistic regression	77.51%
Multilayer perceptron	81.30%
Extra trees	**82.93%**
Random forest	**82.93%**
Bagging	81.03%
Stacking	80.22%

Table: Test set accuracy of various classifiers on the All Vowels dataset

Neural approach

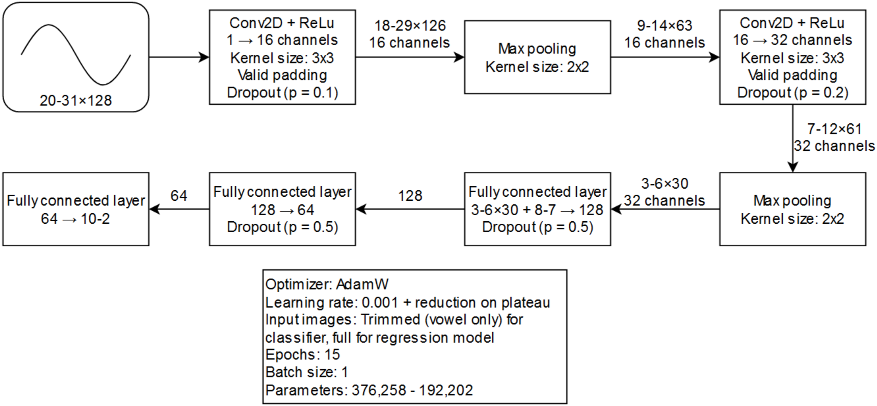


Figure: Best architecture for the neural classifier (left) and vowel extractor (right)

Neural approach

Vowel extractor:

- Total MSE: 0.69371
- 17/356 noticeable errors
- 2 significant ones

Neural classifier:

- Accuracy: 94.5946%

Human Evaluation

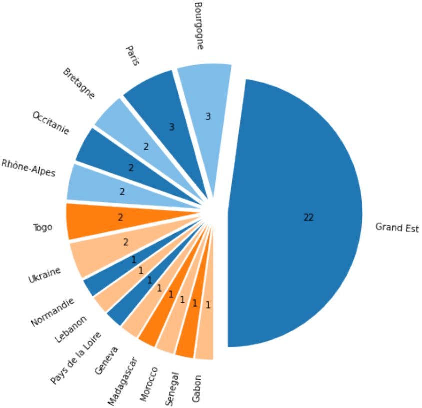


Figure: The origin of the 46 speakers (27 f, 19 m) who tested the system.

Overall results

Accuracy	Male speakers	Female speakers	All speakers
Neural model	50.00%	53.59%	51.56%
Linguistic model	**60.39%**	**77.69%**	**67.89%**

Table: Accuracy of the models depending on the speaker's gender.

Per vowel accuracy

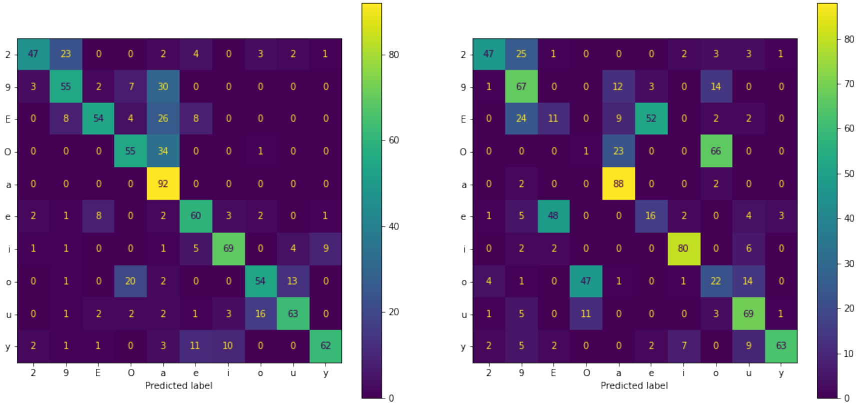


Figure: Confusion matrix of the linguistic (left) and the neural (right) model.

Conclusion and Future work

Conclusion

Neural model:
Better in theory (94.59%), worse in real life (51.56%).

Linguistic model:
Better in real life (67.89 %), worse in theory (82.93%).

Problems of neural model:

- Not noise robust
- Trained on homogeneous input

Problems with linguistic model:

- Performance cap?
- Doesn't leverage confidence much

Future work

Improvements:

- Better sound level detection
- More training data (e.g. from app users)
- Improve NN robustness (data augmentation with noise)
- Gender input $\rightarrow$ types of voices

Extensions:

- Include nasal vowels
- Custom feedback based on the user's native language
- Include consonants
- Add more varieties of French

Merci!

/mɛʁsi/

References

[1] Jessica S Miller. Teaching french pronunciation with phonetics in a college-level beginner french course. *NECTFL Review*, 69:47–68, 2012.

[2] Nancy F Chen and Haizhou Li. Computer-assisted pronunciation training: From pronunciation scoring towards spoken language learning. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–7. IEEE, 2016.

[3] Bernard Rochet. Douglas c. walker. pronunciation of canadian french. ottawa: University of ottawa press. 1984. pp. xxii 185. $15.00 (softcover). *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 32(1):101–107, 1987.

[4] Zsuzsanna Fagyal, Douglas Kibbee, and Frederic Jenkins. *French: A Linguistic Introduction*. Cambridge University Press, 2006.

Accuracy per vowel

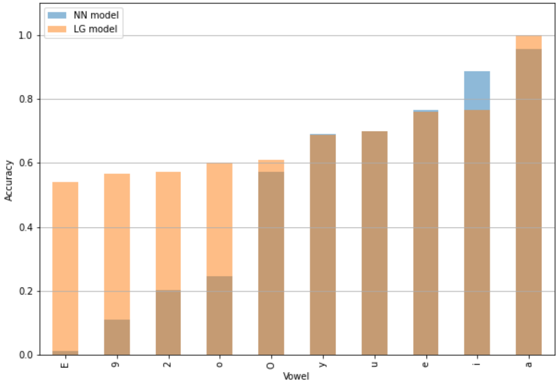


Figure: Accuracy of the neural (NN) and linguistic (LG) models for each true vowel in the dataset.

Model confidence

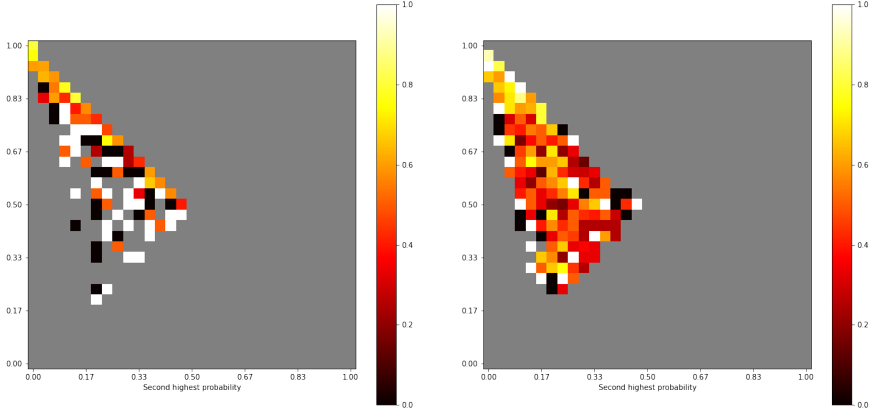


Figure: Accuracy of the linguistic (left) and neural (right) models depending on the value of the highest and second-highest probability returned.