Vowel Tuner

Soklong HIM Nora LINDVALL Maxime MÉLOUX
Jorge VASQUEZ-MERCADO

NLP M2

Software Project
Nov. 7, 2022

UNIVERSITÉ DE LORRAINE

Outline

1. Introduction
2. Rule-based approach
3. First experiments and results
4. Deep learning approach
5. Timeline

Our idea

Main goal

Help language learners improve their pronunciation of French oral vowels

Two approaches

- Rule-based approach
- Deep learning approach

Rule-based approach

Plan

- Extract formants from wav file
- Compare vowels
- Provide feedback

Extracting formants

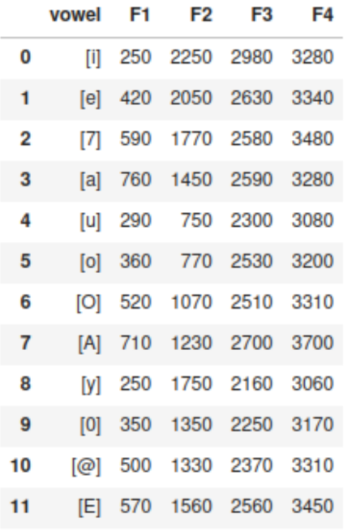
For extracting formants we use the python library **Parselmouth**[1]:

- take the audio file as input
- computes the list of each formant
- computes the average value of each formant

[1] https://parselmouth.readthedocs.io

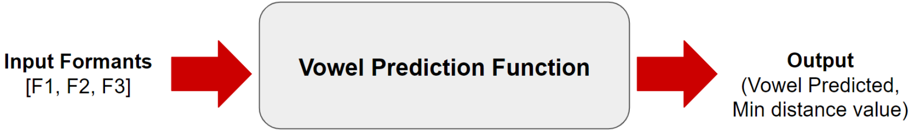
Prediction of the vowel

Vowel formants defined by: *Reconnaissance de phonèmes par analyse formantique dans le cas de transitions voyelle-consonne.*



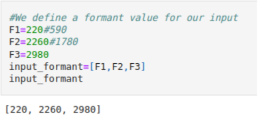
	vowel	F1	F2	F3	F4
0	[i]	250	2250	2980	3280
1	[e]	420	2050	2630	3340
2	[7]	590	1770	2580	3480
3	[a]	760	1450	2590	3280
4	[u]	290	750	2300	3080
5	[o]	360	770	2530	3200
6	[O]	520	1070	2510	3310
7	[A]	710	1230	2700	3700
8	[y]	250	1750	2160	3060
9	[0]	350	1350	2250	3170
10	[@]	500	1330	2370	3310
11	[E]	570	1560	2560	3450

Vowel Prediction Function



Example

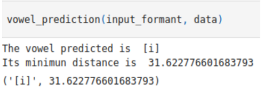
Input Formants



```
#We define a formant value for our input
F1=220#590
F2=2260#1780
F3=2980
input_formant=[F1,F2,F3]
input_formant
```

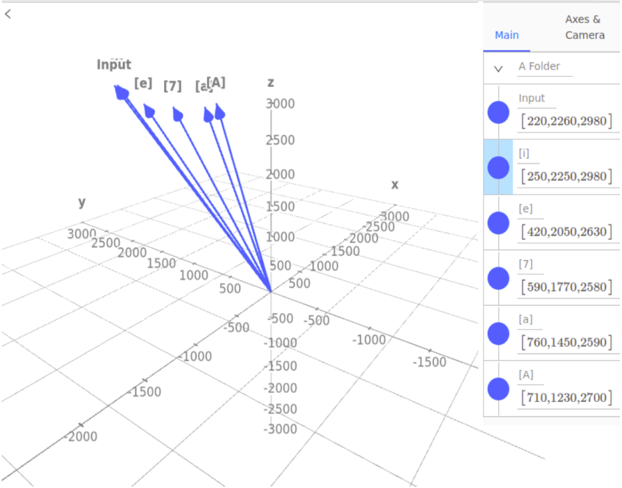
```
[220, 2260, 2980]
```

Executing the function



```
vowel_prediction(input_formant, data)
```

```
The vowel predicted is  [i]
Its minimun distance is  31.622776601683793
('[i]', 31.622776601683793)
```



Input	[220,2260,2980]
[i]	[250,2250,2980]
[e]	[420,2050,2630]
[7]	[590,1770,2580]
[a]	[760,1450,2590]
[A]	[710,1230,2700]

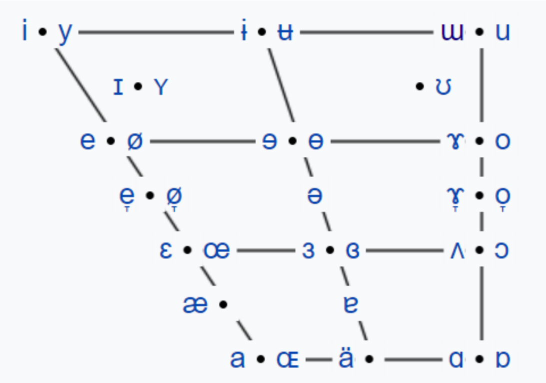
Feedback

- Openness score (int)
- Frontness score (int)
- Rounding (boolean)

”Close your mouth”

”Move your tongue forward!”

”Round your lips!”



Corpus 1

InterFra corpus (Inge Bartning and Fanny Forsberg Lundell), available at https://spraakbanken.gu.se/en/resources/interfra

- 105 hours of L2 French, 4 hours of L1 French
- Transcribed and annotated with Penn POS tags
- Many speaker groups

Corpus 1

Age	French level	L1s
13	3 years	Swedish, Russian, (English)
19-25	beginner	Swedish, Estonian, Spanish, Latvian
19-25	3.5-6 years (4 terms)	Swedish
19-25	3.5-6 years	Swedish
25-35	7-8 years, future teacher	Swedish
25-30	9-10 years	Swedish
25-30	10+ years in France	Swedish
40-50	15-35 years in France	Swedish, (Italian)
19-25	native, northern France	French, (Swedish, Portuguese, Italian)
25-30	native	French, (Spanish)
40-50	native	French, (Italian)
20-35	native	French

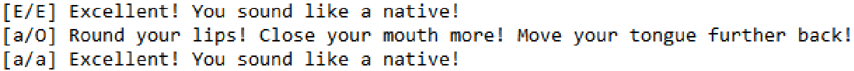
Corpus 1

- 4 speakers selected: 2 non-native (M Swedish/F Estonian), 2 native (M/F)
- 50 vowels or 30 seconds of vowels
- Annotated with left and right phonemic context
- Annotated with perceived vowel
- Result: 225 vowels

Results (corpus 1)

Subset	2 formants	3 formants	4 formants
Native speakers	0.120	0.133	**0.157**
Non-native speakers	0.170	**0.205**	0.114
Female speakers	0.178	**0.208**	0.168
Male speakers	0.100	**0.114**	0.086
Overall	0.146	**0.170**	0.135

Table: Accuracy between the detected and perceived vowels in the InterFra sub-corpus



```
[E/E] Excellent! You sound like a native!
[a/O] Round your lips! Close your mouth more! Move your tongue further back!
[a/a] Excellent! You sound like a native!
```

Corpus 2

Better than random chance! ($1/12 \approx 0.083$)
It seems using 3 formants is the best.

But...

- Reference vowels are for male speakers
- Context, speed
- Difficulty and subjectivity of annotation

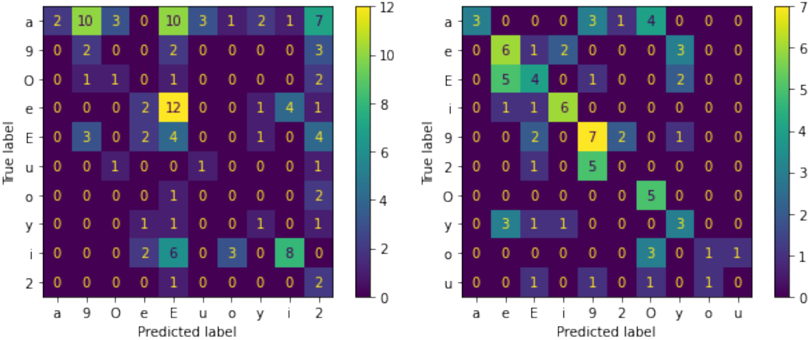
→ Back to the experimental corpus

Results (corpus 2)

Subset	2 formants	3 formants	4 formants
Native speakers	0.312	**0.359**	0.344
Non-native speakers	**0.359**	0.256	0.282
Female speakers	**0.346**	0.192	0.038
Male speakers	0.325	0.364	**0.416**
Overall	**0.330**	0.320	0.320

Table: Accuracy between the detected and perceived vowels in the experimental corpus

Analysis



	a	9	O	e	E	u	o	y	i	2
a	2	10	3	0	10	3	1	2	1	7
9	0	2	0	0	2	0	0	0	0	3
O	0	1	1	0	1	0	0	0	0	2
e	0	0	0	2	12	0	0	1	4	1
E	0	3	0	2	4	0	0	1	0	4
u	0	0	1	0	0	1	0	0	0	1
o	0	0	0	0	1	0	0	0	0	2
y	0	0	0	1	1	0	0	1	0	1
i	0	0	0	2	6	0	3	0	8	0
2	0	0	0	0	1	0	0	0	0	2

	a	e	E	i	9	2	O	y	o	u
a	3	0	0	0	3	1	4	0	0	0
e	0	6	1	2	0	0	0	3	0	0
E	0	5	4	0	1	0	0	2	0	0
i	0	1	1	6	0	0	0	0	0	0
9	0	0	2	0	7	2	0	1	0	0
2	0	0	1	0	5	0	0	0	0	0
O	0	0	0	0	0	0	5	0	0	0
y	0	3	1	1	0	0	0	3	0	0
o	0	0	0	0	0	0	3	0	1	1
u	0	0	1	0	1	0	1	0	1	0

Figure: Confusion matrices on the InterFra sub-corpus and the experimental corpus

Analysis

A few discussion points:

- Significantly better!
- How many formants?
- Pertinence of the reference vowels
- Still a single annotator, highly subjective

What can be improved?

- The metric
- The input (more features)
- The annotations
- The method

Deep learning approach

First draft:

- Corpus annotated with perceived vowels
- Train model to recognize vowel (classification)
- Compare output vowel to target vowel

Issue

Sometimes /i/ and /i/ are different

Deep learning approach

Second draft:

- Find corpus annotated with perceived vowel + fluency score
- Train model to recognize vowel
- Add classifier for fluency score
- Compare output vowel to target vowel
- If phoneme is correct, give fluency feedback

Issue

Corpus annotated with *perceived vowels?*

Corpus creation

- Size: 28 individuals
  - 50% native, 50% non-native
  - 50% male, 50% female
  - preferably native speakers from the same region
- Recording: real-life conditions
- Annotation: by French native speakers
  Did you hear 'doux' or 'du'?
  Did you hear 'o' as in 'mot' or as in 'mort'?

Plan

Task partition

- Data collection - All
- Corpus annotation - Maxime
- Provide *good* feedback - Nora
- Create/train model - Jorge
- Create interface - Soklong

Provisional timeline

- Corpus complete by Nov 30
- Feedback plan ready by Nov 30
- Model trained and evaluated by Dec 7
- Interface ready by Jan 13

Mitigation plan

If it doesn't work out?
Abandon deep learning and perfect rule-based approach

Thank you!

Questions? Feedback?