Vowel Tuner

Soklong HIM Nora LINDVALL Maxime MÉLOUX
Jorge VASQUEZ-MERCADO

NLP M2

Software Project
Jan. 23, 2023

Outline

1. Recap
2. Experimental set-up
3. Dataset
4. Results
5. Retrospective

Plan

1. Recap
2. Experimental set-up
3. Dataset
4. Results
5. Retrospective

Two approaches

Linguistic approach:
Formants, gender and phonemic context

$\rightarrow$ 82.93% in theory

Neural approach:
Mel-spectrogram of the vowel, gender and and phonemic context

$\rightarrow$ 94.59% in theory

Which one performs better in real-life conditions?

Plan

1 Recap

2 Experimental set-up

3 Dataset

4 Results

5 Retrospective

Logistics

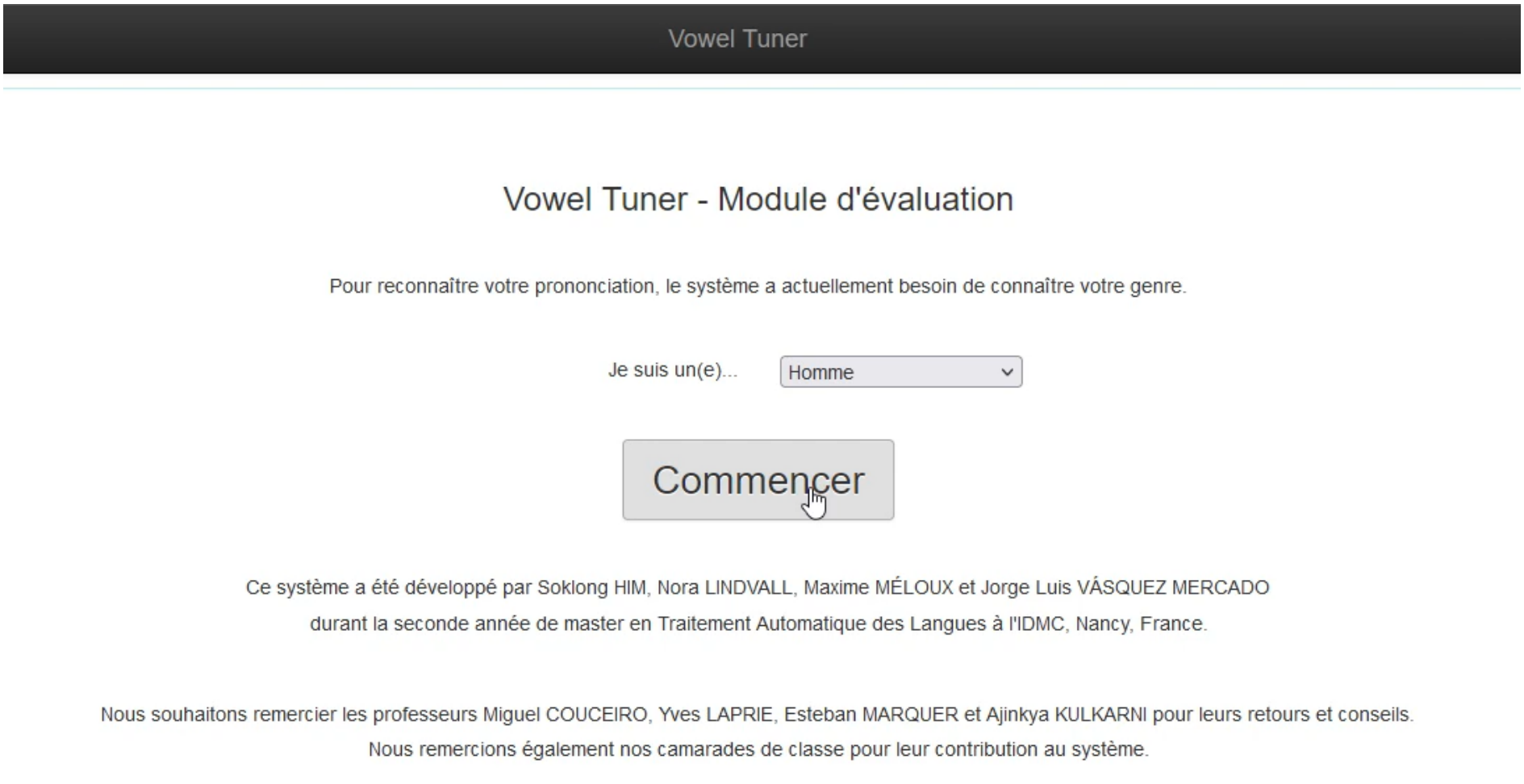
We decided to test our models on real users.

- Created an evaluation module
- Set up a recording station
- Real-life conditions: Noisy room with other speakers
- Put up posters around IDMC
- Incentivized with candy

The experiment

- Users recorded 10-30 monosyllabic words.
  la, *si*, *serre*, *mot*, *tort*, *les*, *peu*, *meurt*
- Performance of neural and linguistic models is compared
- Vowel extractor is always used (affects both models equally)
- We did NOT keep the recordings
- Only user *gender* and *place of origin* (voluntary inputs)

Demo


Vowel Tuner
Vowel Tuner - Module d'évaluation
Pour reconnaître votre prononciation, le système a actuellement besoin de connaître votre genre.
Je suis un(e)...
Homme
Commencer
Ce système a été développé par Soklong HIM, Nora LINDVALL, Maxime MÉLOUX et Jorge Luis VÁSQUEZ MERCADO
durant la seconde année de master en Traitement Automatique des Langues à l'IDMC, Nancy, France.
Nous souhaitons remercier les professeurs Miguel COUCEIRO, Yves LAPRIE, Esteban MARQUER et Ajinkya KULKARNI pour leurs retours et conseils.
Nous remercions également nos camarades de classe pour leur contribution au système.

Plan

1 Recap

2 Experimental set-up

3 Dataset

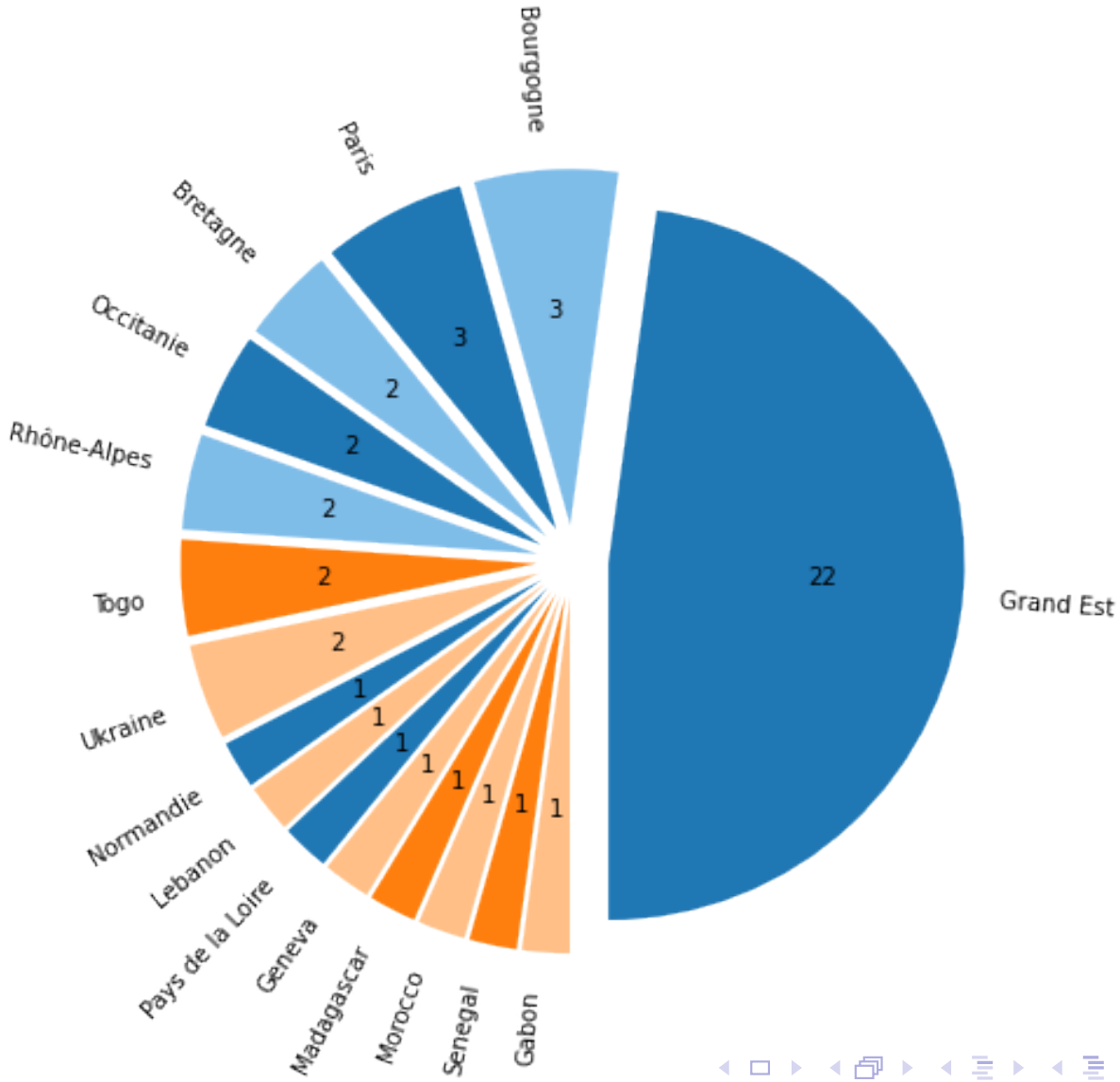
4 Results

5 Retrospective

Dataset

- 46 users partook (27 female, 19 male)
- Three non-native speakers (Lebanon, Ukraine)
- Most people recorded 20 words
- Total dataset of **900** words

Origin distribution



Plan

1 Recap

2 Experimental set-up

3 Dataset

4 Results

5 Retrospective

Accuracy per vowel

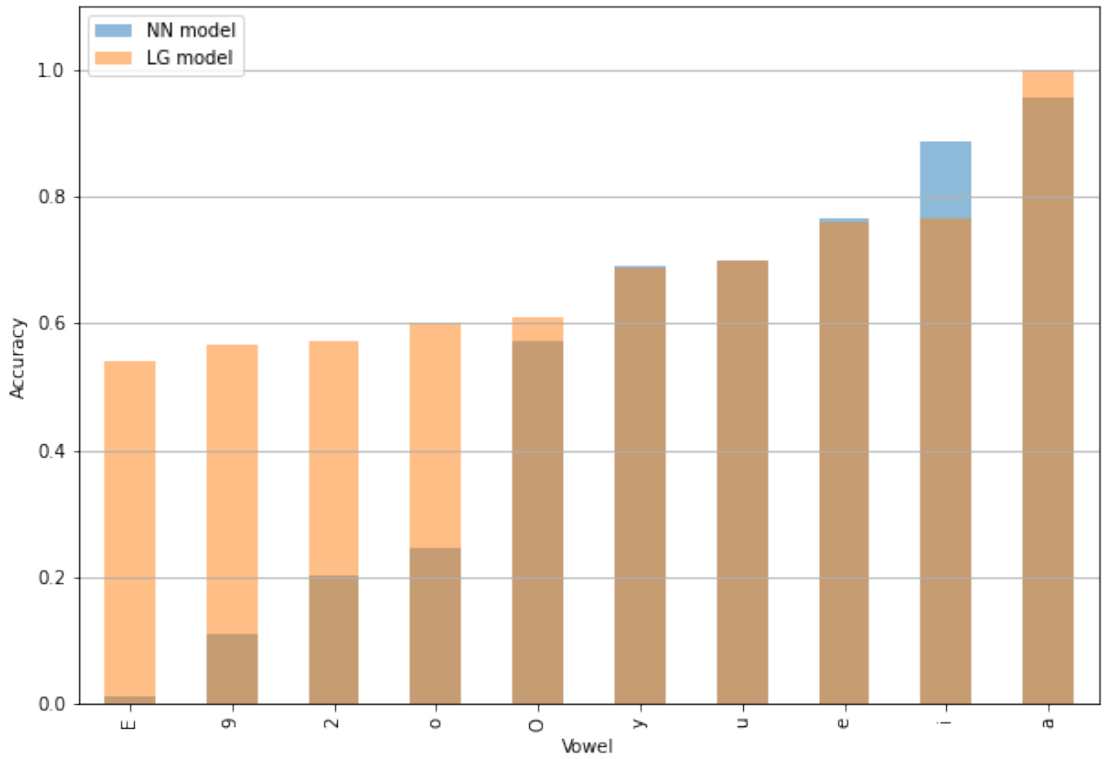


Figure: Accuracy of the neural (NN) and linguistic (LG) models for each true vowel in the dataset.

Accuracy per vowel

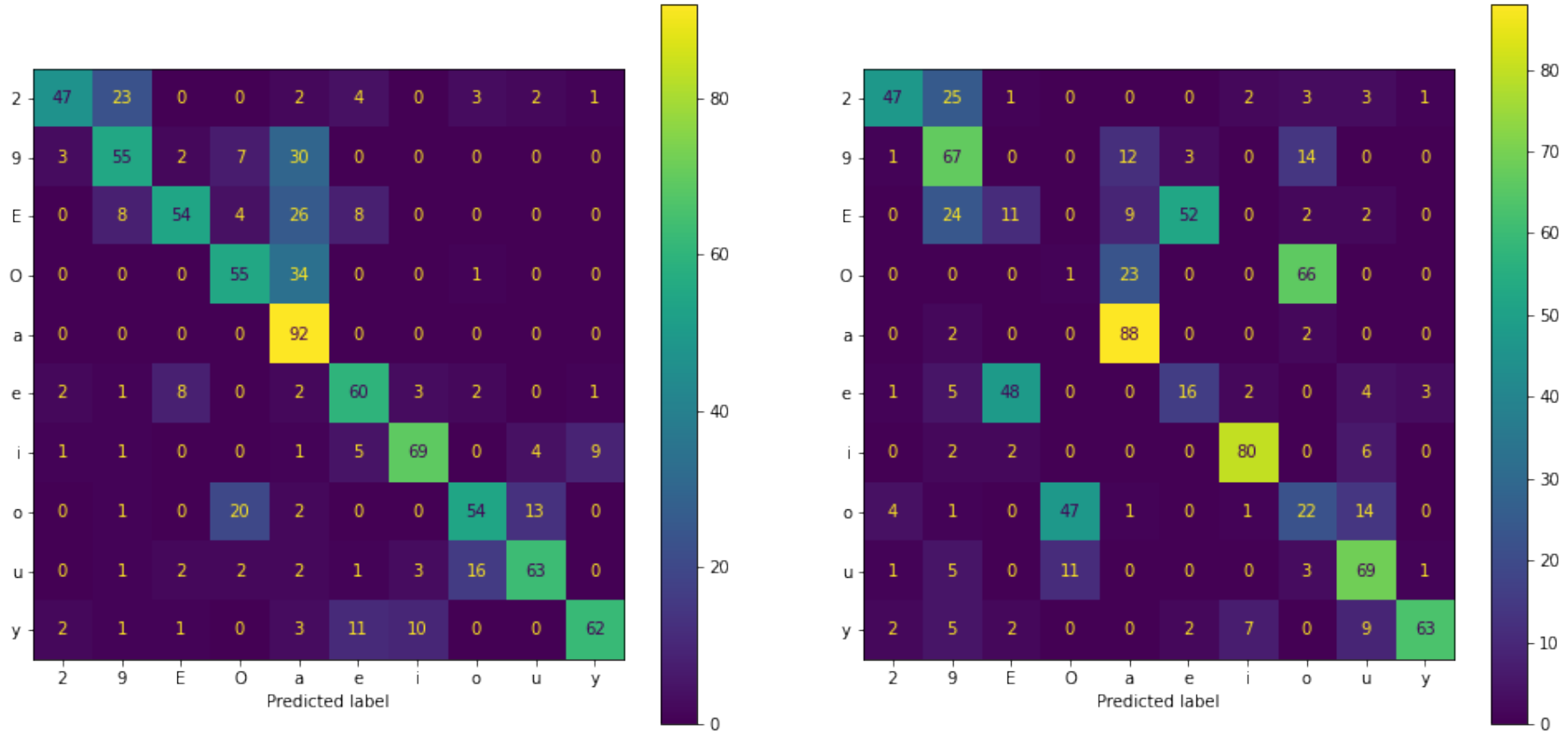


Figure: Confusion matrix of the linguistic (left) and neural (right) model.

Speaker distribution

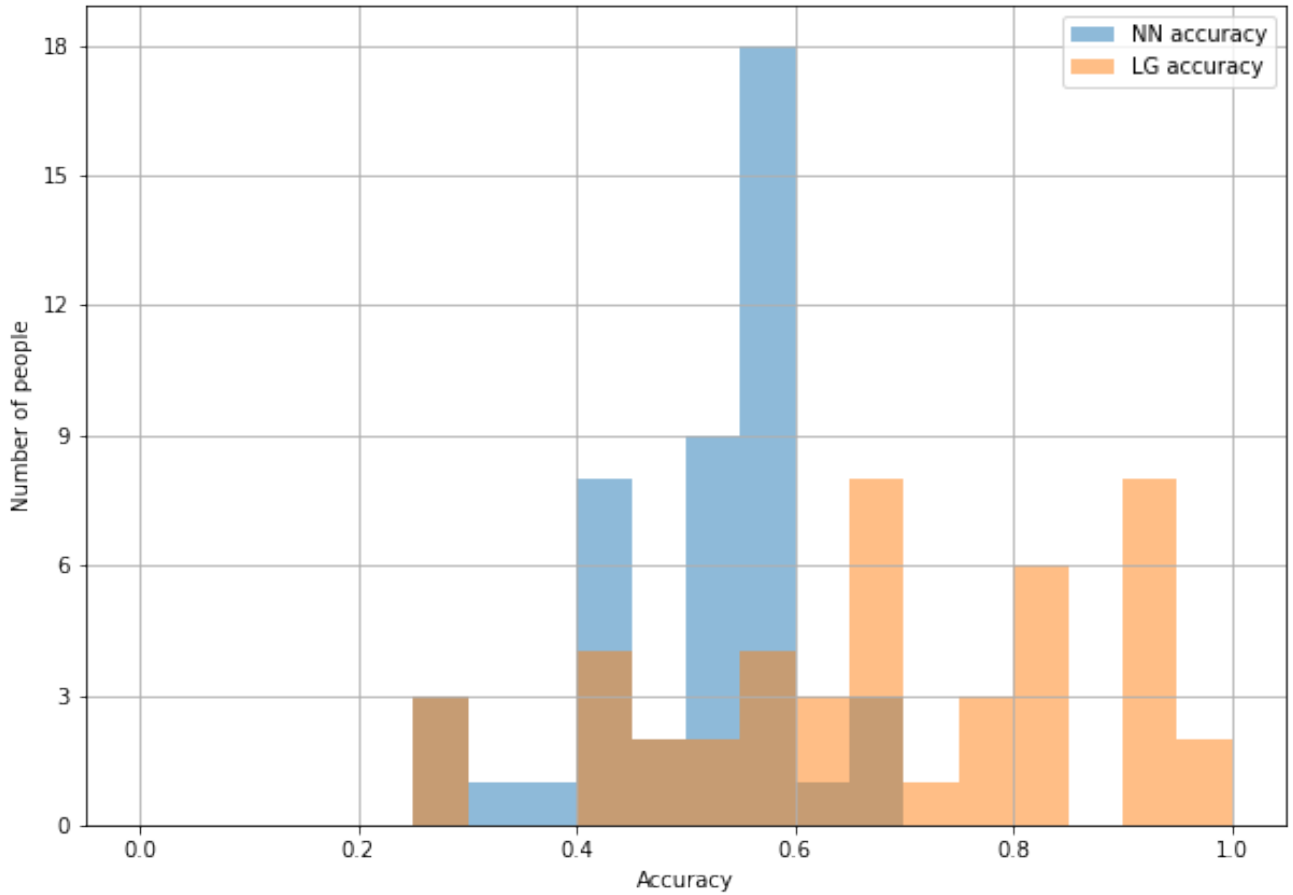


Figure: Distribution of the accuracy of the neural and linguistic models accuracy among speakers.

Model confidence

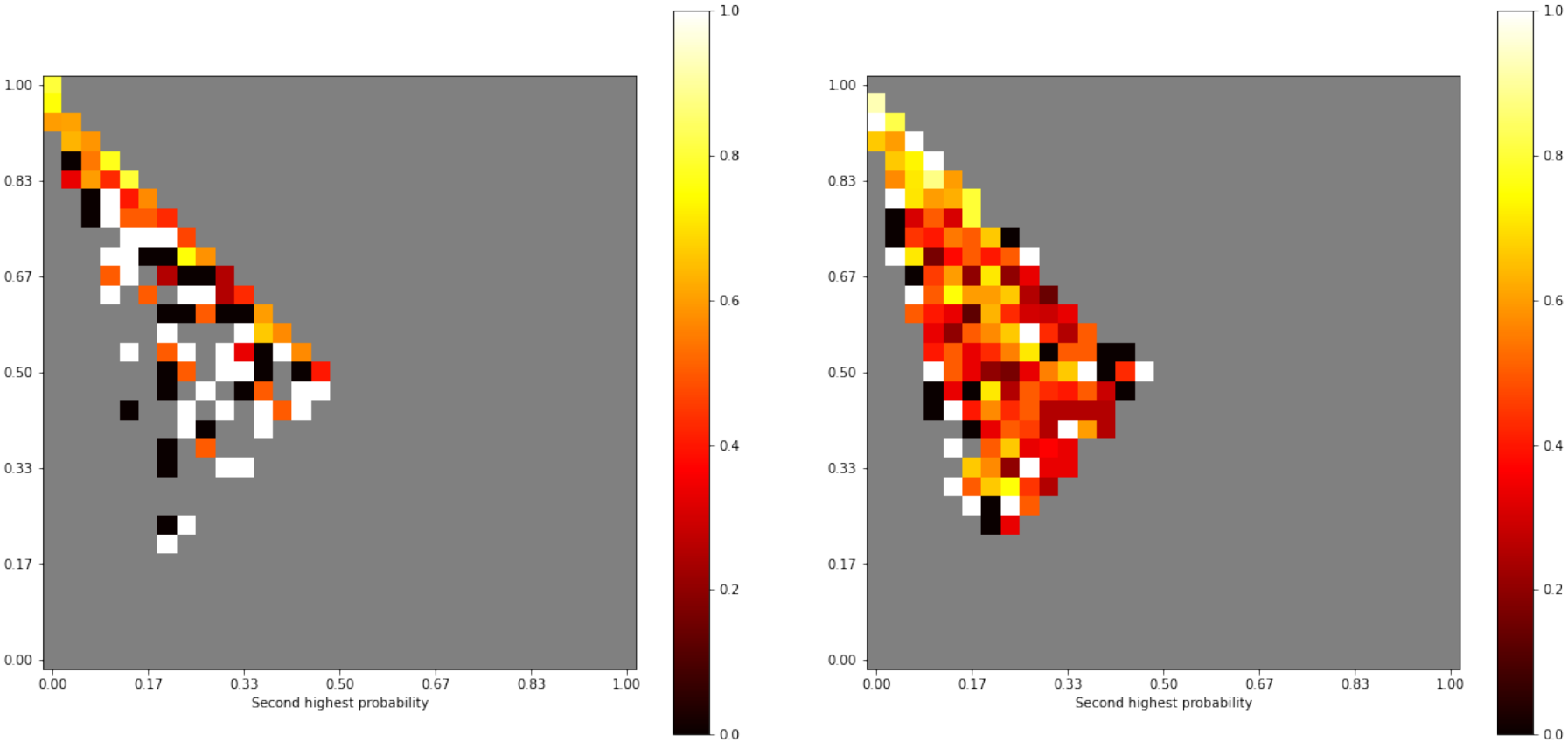


Figure: Accuracy of the linguistic (left) and neural (right) models depending on the value of the highest and second-highest probability returned.

Model confidence

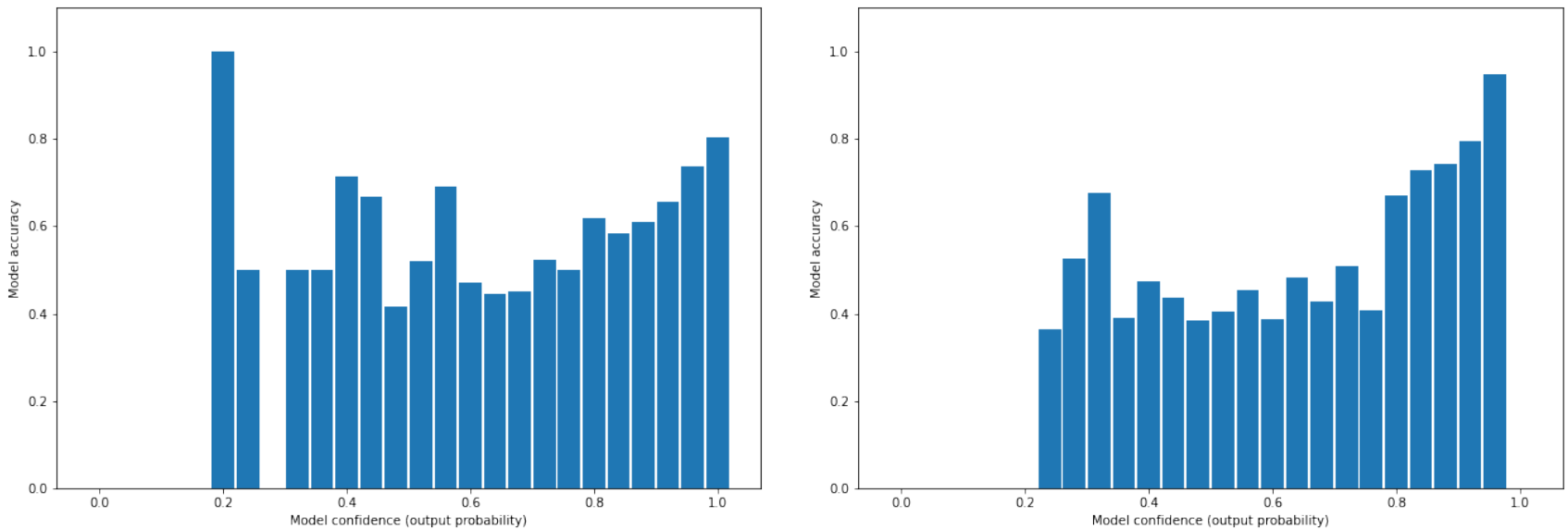


Figure: Accuracy of the linguistic (left) and neural (right) models depending on the value of the highest probability returned.

Results

So the winner is....

The Linguistic Model

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.

"Effective" accuracy: much higher?

Plan

1 Recap

2 Experimental set-up

3 Dataset

4 Results

5 Retrospective

Conclusion

Neural model - better in theory, worse in real life. **Why?**

- Not noise robust
- Not enough data

Model confidence corresponds to accuracy.

Linguistic model - worse in theory, better in real life.

- Noise robust
- Performance cap?

Doesn't leverage confidence much.

Future work

The application...

- Satisfactory results for such a short project
- Is it usable yet $\rightarrow$ Maybe?

Easy improvements:

- Better sound level detection
- More training data (e.g. from app users)
- Improve NN robustness (data augmentation with noise)

Extensions

Our plans haven't changed:

- Remove gender input if possible
- Nasal vowels (easy)
- Custom feedback based on the user's native language (long)
- Consonants (hard, especially for feedback)
- More varieties of French (hard)

Thank you

Thank you to everyone who helped us evaluate our model!

That's all Folks!