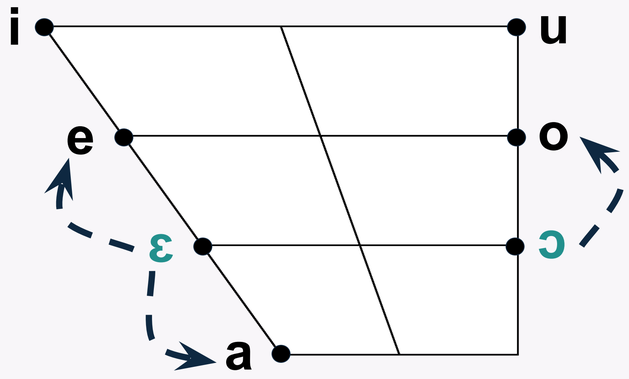
# Small Neural Networks as Models of Cross-Linguistic Speech Perception

Annika Shankwitz
University of Maryland – ashankwi@umd.edu

## TL;DR

Spanish has five vowels [i e a o u]; Catalan has seven vowels [i e ɛ a ɔ o u]




vowels common to Spanish and Catalan

vowels specific to Catalan

**Difficult Catalan vowel contrasts for Spanish -dominant adult listeners** (Amengual, 2016)

- /ɛ/~/e/
- /ɛ/~/a/
- /ɔ/~/o/

**L1 Filter** – Given a listener's native language (L1), certain non-native speech sounds are harder to distinguish than others

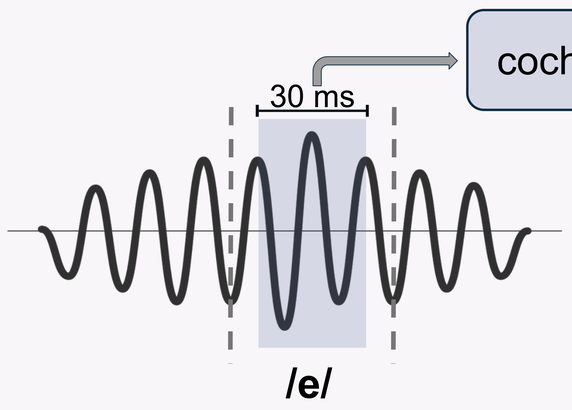
↪ Unclear what combination of *model size* and *learning style* best mimics the L1 filter

**Small supervised neural networks trained on realistic input *can mimic* the L1 filter**

↪ Models trained on Spanish classify Catalan /ɛ/ as /e/ or /a/, and Catalan /ɔ/ as /o/

## Data and Preprocessing

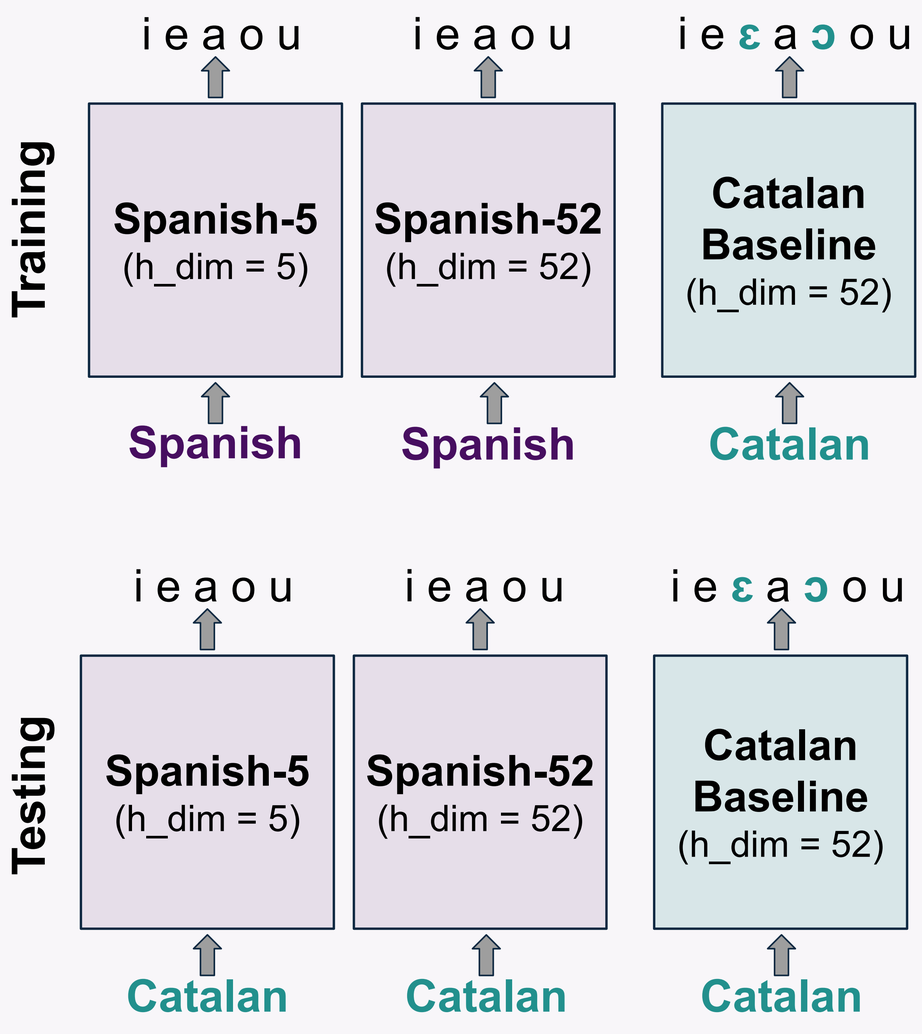
- Spanish Corpora: Multilingual Librispeech (918 hrs)
- Catalan Corpora: Crowdsourced high-quality Catalan speech data set (9.25 hrs)

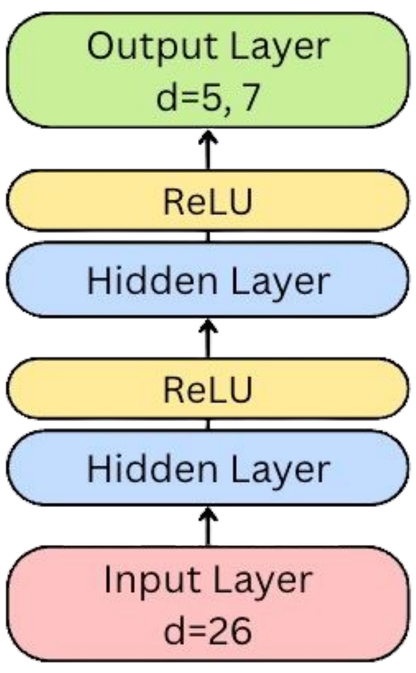



| | Spanish | | | Catalan | |
|---|---|---|---|---|---|
| | train | test | dev | train | test |
| total vowels | 18,750 | 5,000 | 5,000 | 5,880 | 1,680 |

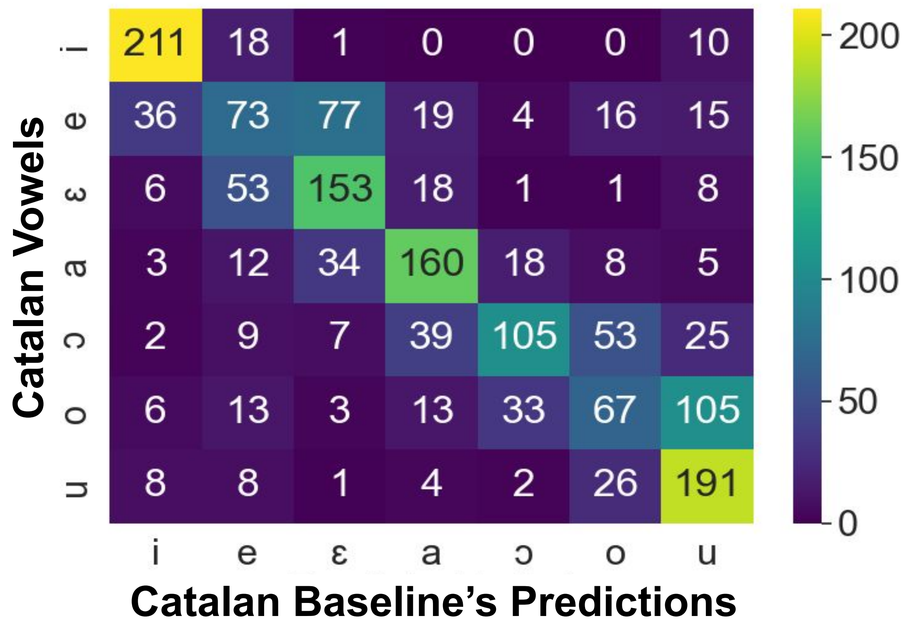
## Model Architecture and Training

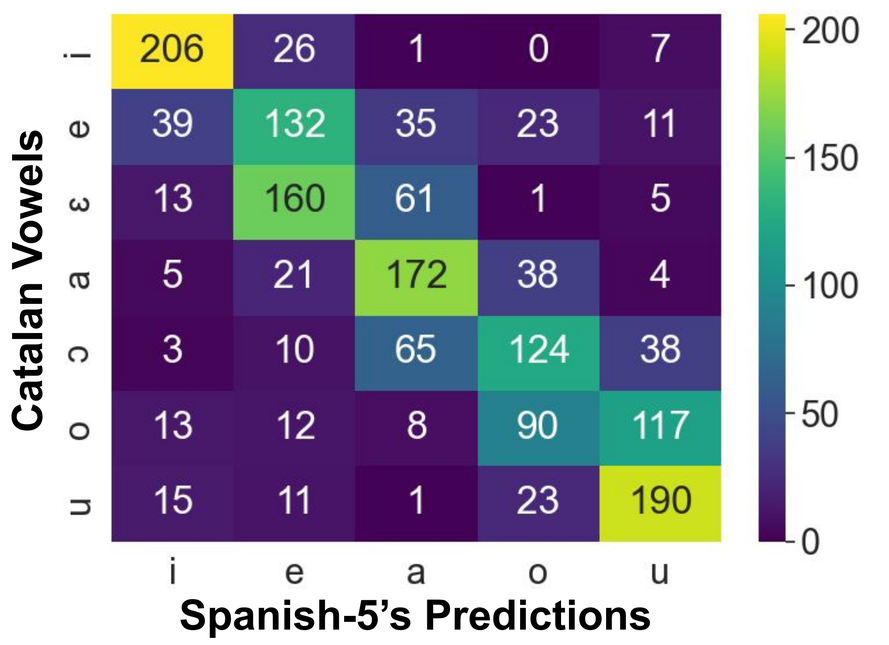
Four-layer feedforward neural networks (two hidden layers)

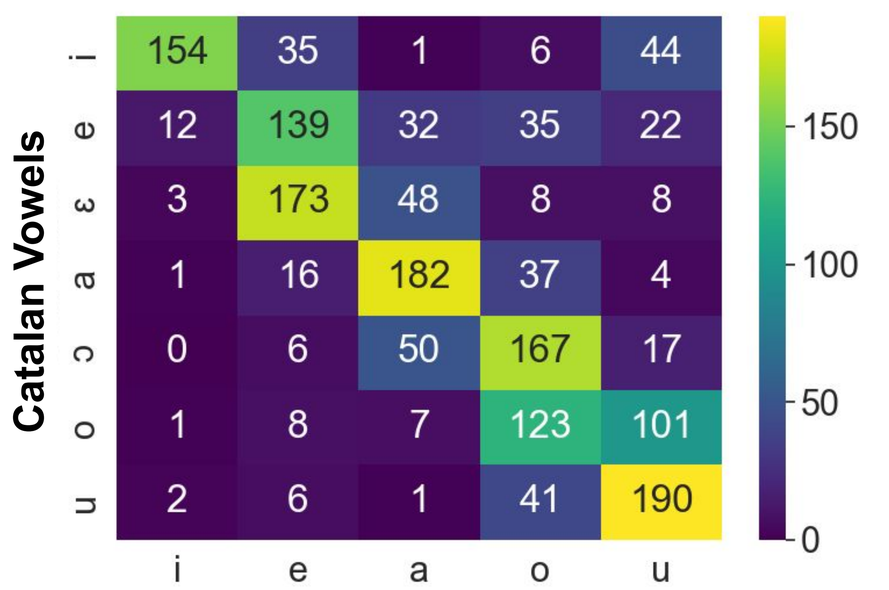



***Spanish models should classify /ɛ/ as /e/ or /a/ and /ɔ/ as /o/***




| | /ɛ/ as /e/ | /ɛ/ as /a/ | /ɔ/ as /o/ |
|---|---|---|---|
| Spanish-5 | 67% | 25% | 52% |
| Spanish-52 | 72% | 20% | 70% |
| Baseline | 22% | 8% | 22% |

**/ɛ/ → /e/ or /a/ ✓**

**/ɔ/ → /o/ ✓**

**Using just 30 ms of speech, small supervised neural networks can mimic the L1 filter, and thus show promise as compact models of cross-linguistic speech perception**

**References**

- Mark Amengual. 2016. The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. International Journal of Bilingualism, 20(2):133–152.
- Oddur Kjartansson, Alexander Gutkin, Alena Butryna, Isin Demirsahin, and Clara Rivera. 2020. Open-source high quality speech datasets for Basque, Catalan and Galician. In Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages and Collaboration and Computing for Under-Resourced Languages, pages 21–27, Marseille, France. European Language Resources association.
- Vineel Pratap, Qiantong Xu, Anuroop Sriram, Gabriel Synnaeve, and Ronan Collobert. 2020. MLS: A large-scale multilingual dataset for speech research. In Interspeech 2020, pages 2757–2761. ISCA.