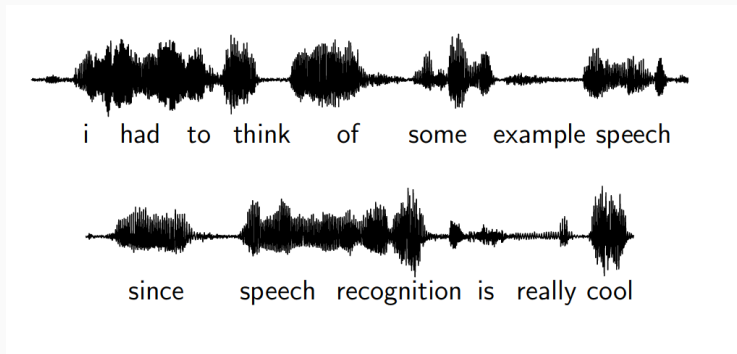


Removing Nuisance Variables from Acoustic Word Embeddings

Lisa van Staden

Low-Resource Speech and Language Processing

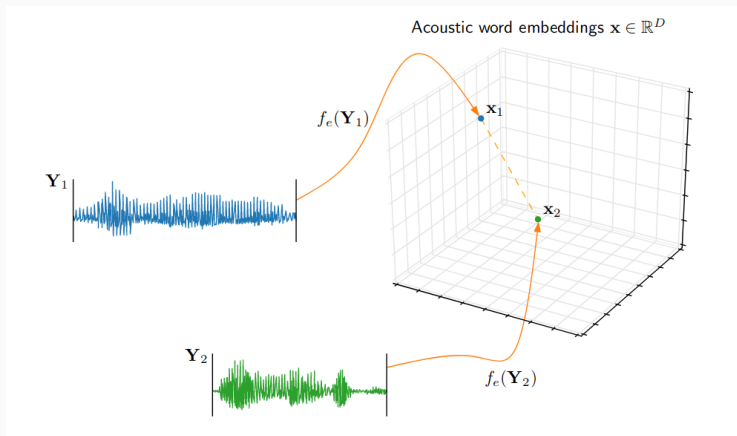


Popular methods for speech processing rely on transcribed speech. Obtaining transcriptions is expensive and not always possible.

We don't always need to predict text labels:

- Query-by-Example Search: search speech using speech.
- Tasks need speech segments to be compared.

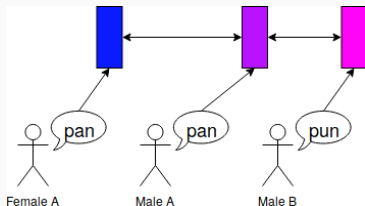
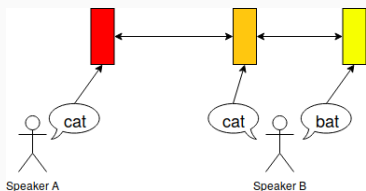
Acoustic Word Embeddings



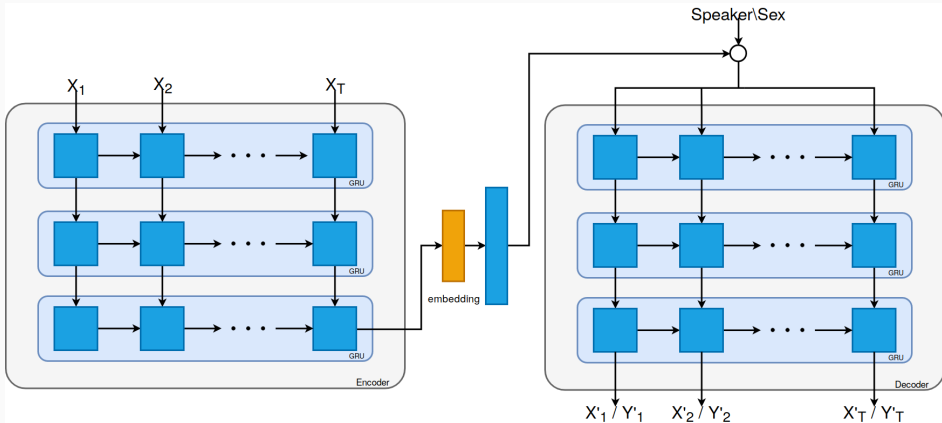
We want to map speech to these representation without using labels.

Nuisance Variables: Speaker and Sex

Acoustic properties of speech from different speakers/sexes differ.
We want embeddings to be robust.



Current Models



What's Next

- Improved models: Disentanglement with adversarial training.
- Using embeddings in downstream tasks.
- Investigate the phonetic information in embeddings.
- Links to language acquisition.