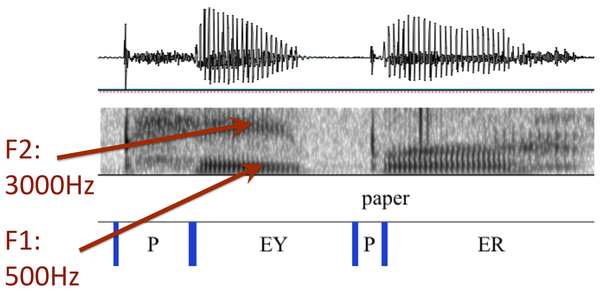
A Web Application for Automated Dialect Analysis

Sravana Reddy & James Stanford, Dartmouth College

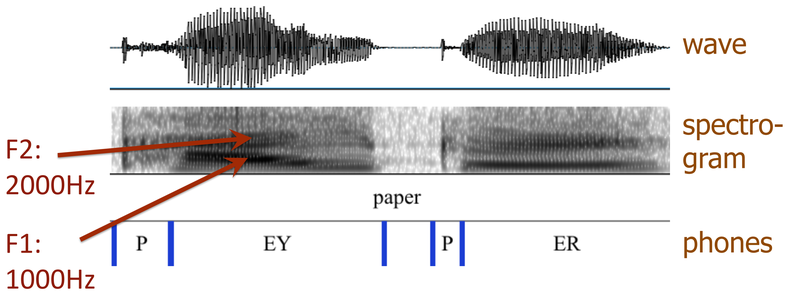
DARLA
darla.dartmouth.edu

Problem: Vowel Formant Extraction

- Socio-phoneticians study accents and social variables.
- Quantify accent with formants (resonance frequencies), F1 & F2.

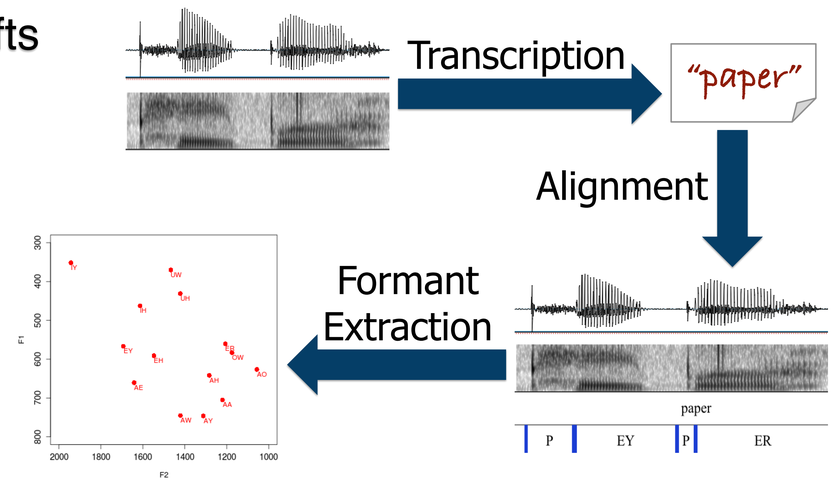


Northern speaker



Southern speaker

- Accents = systematic shifts in formant space.
- Common task: audio → formants of each vowel.



Existing Tools

"Semi-automated" – e.g. FAVE (fave.ling.upenn.edu)

- Alignment: automated with dynamic programming)
- Formant extraction: automated with LPC
- **Transcription: manual**

We now have access to thousands of hours of speech – manual transcription is impossible.

Our Idea

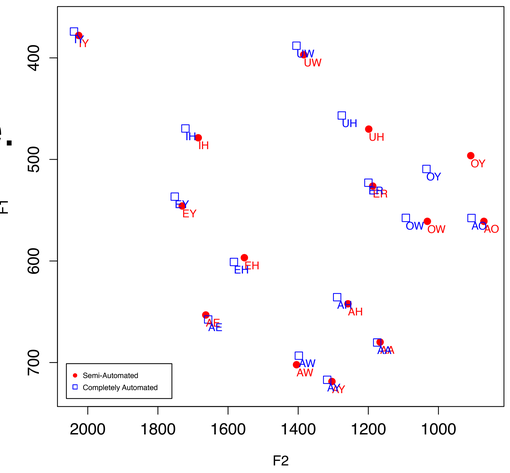
Automate transcription with speech recognition … but isn't speech recognition inaccurate?

Insight: stressed vowels are usually correct

```
REF:  no it's it's wood  turning
HYP:  no it   it   would turn it

REF:  a real dog and cat and all the others
HYP:  a real docking tap and on  the others
```

- Filter out vowels with low acoustic confidence.
- Result: Formants from completely automated system ≅ formants from semi-automated.



Where is Obama from?

Implementation

- Speech recognition with CMU Pocketsphinx
  - Generic English acoustic models trained on LibriSpeech (400 hours), language models on WSJ and Fisher transcripts.
- Alignment and formant extraction with FAVE.
- Web interface accepts files or YouTube links.
- Processing time is about 3x the audio length.