

How Much Data for Stable Formant Values?



UPPSALA
UNIVERSITET

Pipeline for Convergence Detection Based on Read Speech

Kayla Sward¹, Johan Sjons¹, Axel Ekström²

¹Department of Linguistics and Philology, Uppsala University

²Centre for Cultural Evolution, Department of Psychology, Stockholm University

Motivation

- ▶ Studies report average vowel formant values, but rarely specify how much data they use
- ▶ Most measurements come from short, controlled recordings
- ▶ The required amount of speech for stable averages is unclear

Data

LibriVox audiobooks +
King James Bible

47 audiobooks • 3,384 chapters

942 hours
8.8M words

Convergence Methods

- ▶ First Token Within (FTW)
- ▶ Cumulative Sum (CUSUM)
- ▶ Two Sample *t*-Test
- ▶ Confidence Interval (CI) Shrinkage
- ▶ Piecewise Linear Fitting (PWLF)
- ▶ Binary Segmentation (BinSeg)

Results

- ▶ Around 55–220 tokens needed, depending on vowel/formant
- ▶ Female speakers generally stabilized faster than males
- ▶ Vowel differences:
 - ▶ /a/ = slowest
 - ▶ /i/ = fastest
 - ▶ /ə/ = intermediate

Key Finding

Based on convergence across methods and vowels, stable formant means require approximately:

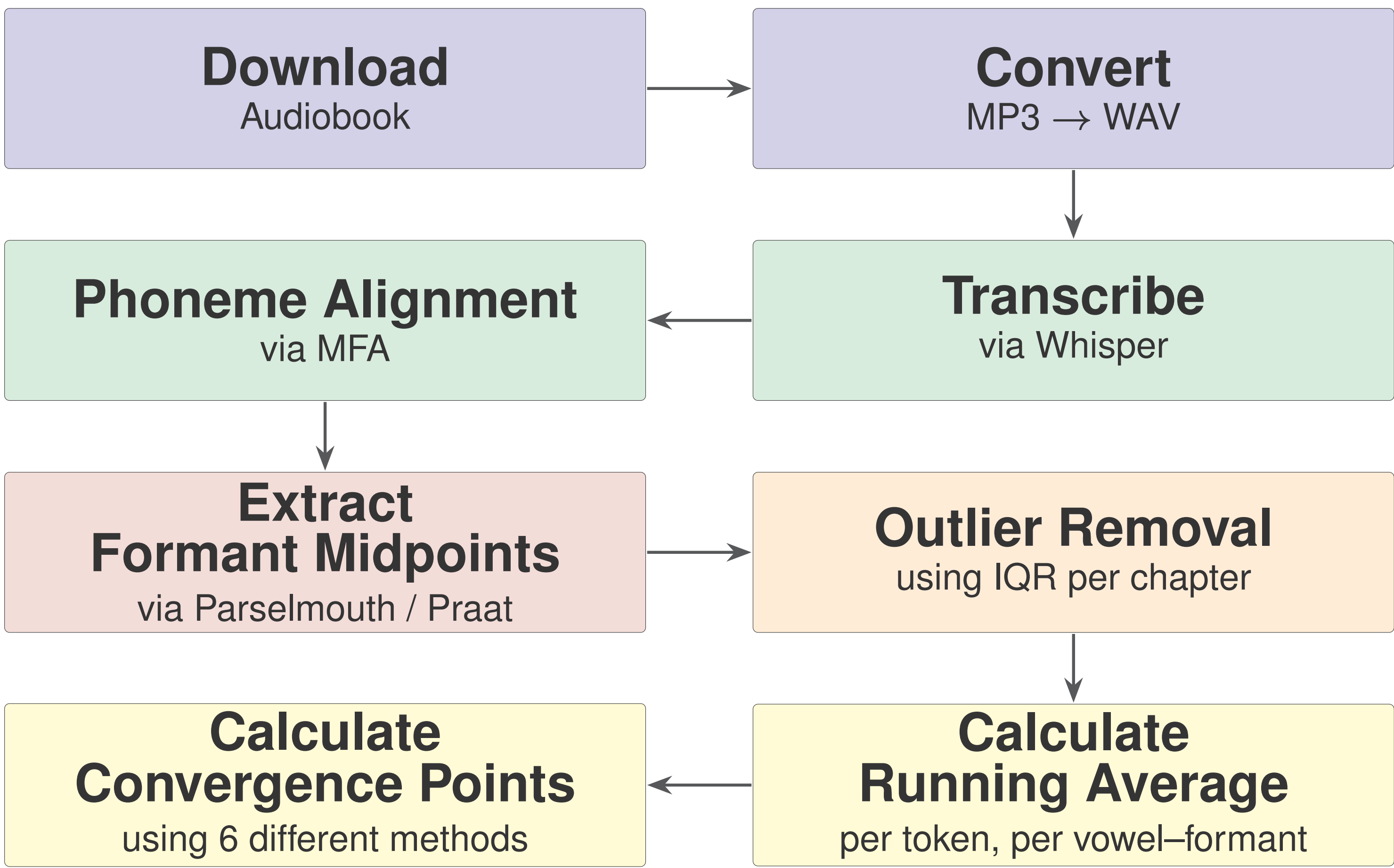
~1,050 words

(~7 minutes of speech)

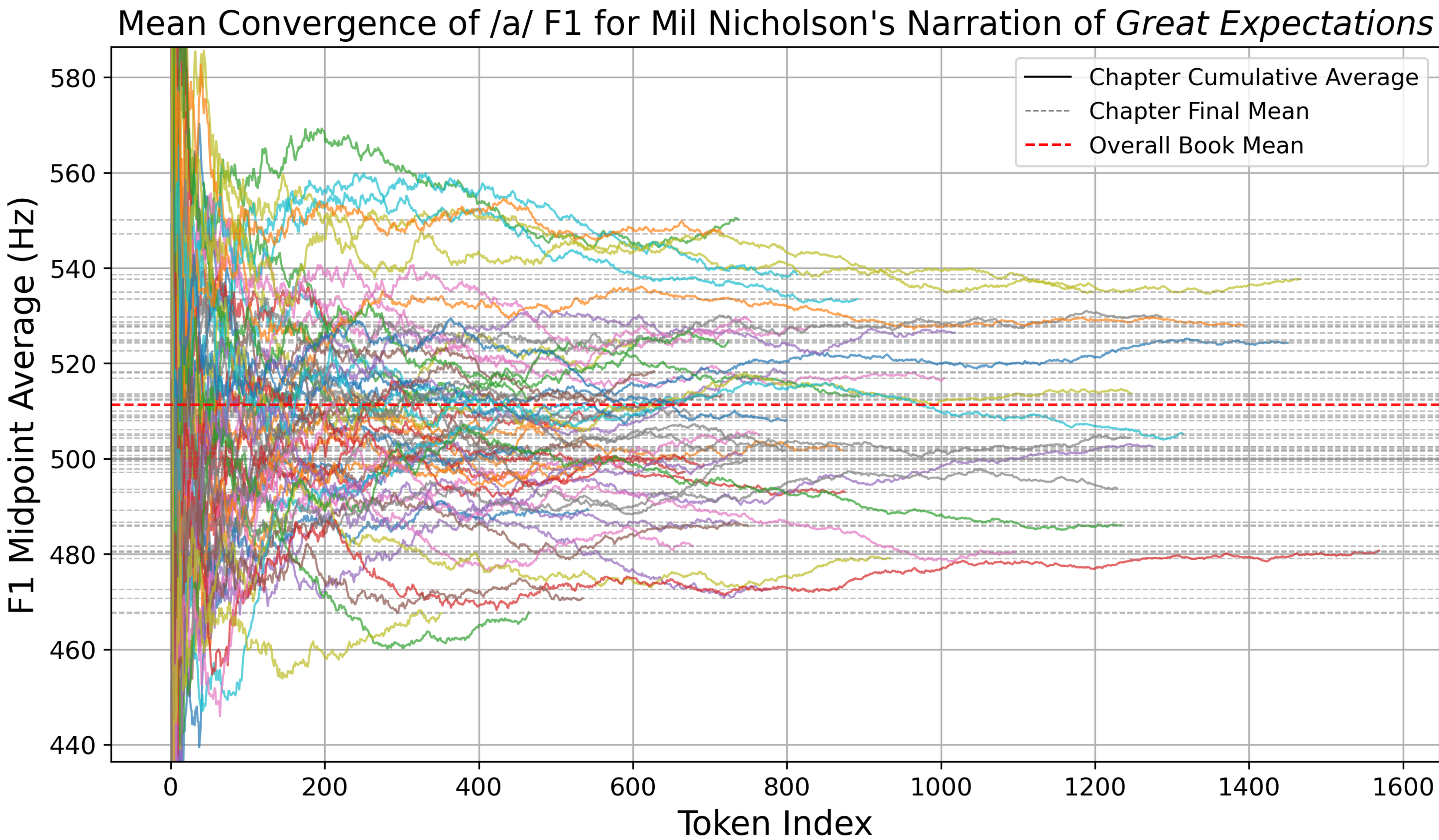
Next Steps

- ▶ Apply the pipeline to conversational & spontaneous speech
- ▶ Evaluate stability across more diverse speakers and speaking styles

Method Pipeline



Convergence Plot Example



Convergence Timeline Example

