USC
UNIVERSITY
OF SOUTHERN
CALIFORNIA

Prosodic Characterization of Reading Styles using Audiobook Corpora

4pSCb32

Michael Proctor, Athanasios Katsamanis

http://sail.usc.edu

162nd Meeting of the ASA
Thursday 06-Nov-11
San Diego, CA

Perception of Read Speech

- native speakers of Germanic Languages have strong intuitions about the felicity of different reading styles: [1]
  - preference for 'spontaneous' speech over read speech
  - preference for human readers over TTS
  - preference for some readers over others
- ***which properties of read speech influence listener preferences and perceptions of felicity?***
- prosodic structures of read speech and spontaneous speech have been shown to differ: ***do prosodic factors contribute to the perception of different reading styles*** as more felicitous?
- can relevant prosodic differences be systematically quantified?

Characterizing Read Speech

- differences in the realization of read speech (c.f. spontaneous): [1-7]
  - higher F0, more F0 variation, more F0 declination
  - lower speech rate + longer pauses
  - longer major tone units
  - less shimmer, less vowel reduction
- less known about the phonetic ***characteristics which differentiate reading styles of different speakers***
- wide variety of metrics have been proposed to capture prosodic variability and stylistic characteristics of speech: [8-10]
  - PVI: pair-wise variability indices
  - ΔV, %V: occurrence, distribution of vocalic intervals
  - ΔC, %C: occurrence, distribution of consonantal intervals
  - VarCoV/C: std. dev of cons/vocalic interval duration/mean
- problems with metric definitions, reproducibility, sample size
- speech style difference ***studies limited by lack of availability of transcribed speech data*** representing the different speech styles under examination

Goals

(i) Examine **listener responses to a range of different readers**:
- to what extent listener preferences are individual or global
- to what extent individual readers are preferred over others

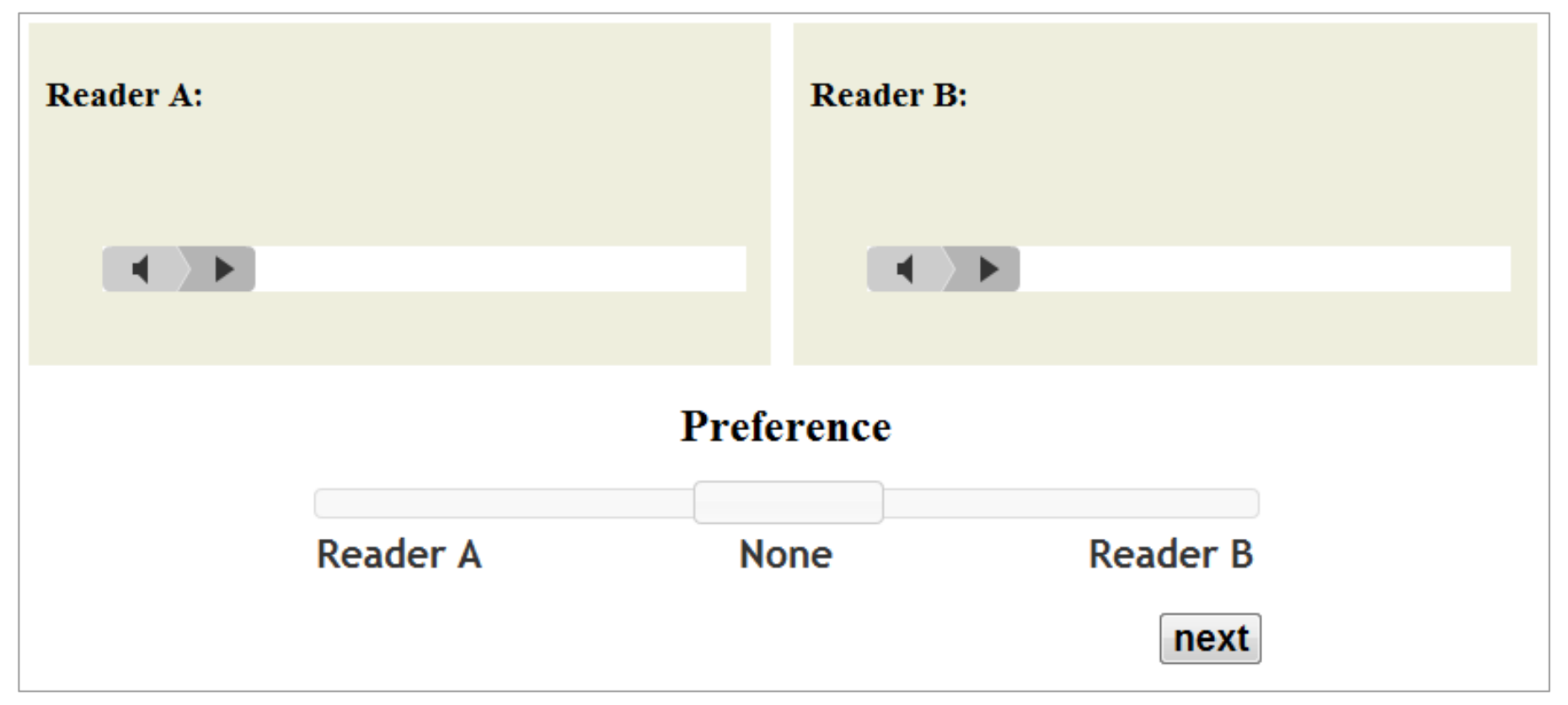
(ii) Examine the **prosodic characteristics of preferred and dispreferred read speech**:
- to what extent does prosody influence perceptions of felicity?
- which metrics best characterize most favored read speech?

(iii) Make use of *underexploited new resources for linguistic research*: **audiobook corpora and companion open-source texts**
- previously pioneered Yuan et al. 2008 and others [11]
- take advantage of massive, freely-available, multi-speaker database containing hours of unanalyzed speech
- rich resource for studying speech styles, prosody, listener responses, & for testing methodologies on large datasets

Method: Listener Preferences

Preferences for reading styles evaluated by asking listeners to evaluate speech samples from different readers, using a head-to-head comparison paradigm:

- ten x 10-second speech samples extracted at random intervals from audio recordings of each reader to be evaluated
- recordings taken from two works of a single author (Jack London) of standard 20th Century American English [12,13]
- auditors: 13 native speakers of General American English
- listeners compared all readers by auditing 3 random samples of each reader, juxtaposed against 3 samples of each other reader
- forced choice/no preference decision task
- hierarchy of readers constructed from cumulative rankings of listener preferences

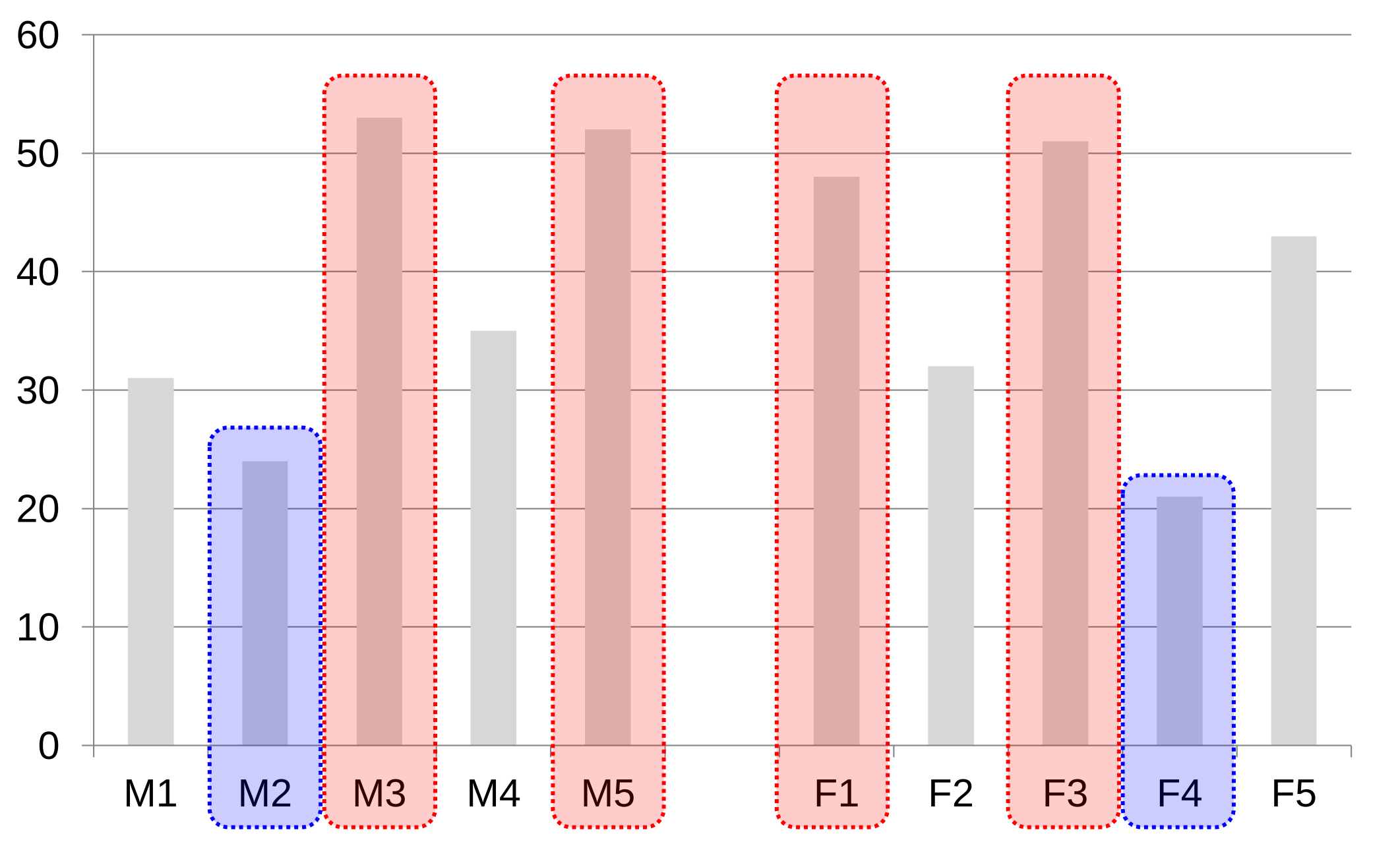


Results: Listener Preferences

- Individual auditor's preferences differ, but overall, clear preferences and dispreferences emerge:

	Male Reader Rankings					Female Reader Rankings				
	1st	2nd	3rd	4th	Last	1st	2nd	3rd	4th	Last
L1	M4	M1	M5	M3	M2	F4	F2	F1	F5	F3
L2	M4	M1	M2	M5	M3	F4	F1	F3	F5	F2
L3	M2	M1	M3	M5	M4	F4	F5	F1	F2	F3
L4	M4	M2	M5	M1	M3	F4	F2	F5	F1	F3
L5	M5	M2	M1	M3	M4	F5	F4	F2	F3	F1
L6	M3	M2	M1	M4	M5	F4	F2	F3	F5	F1
L7	M1	M2	M4	M3	M5	F2	F1	F3	F4	F5
L8	M4	M2	M1	M5	M3	F4	F2	F5	F3	F1
L9	M2	M1	M4	M3	M5	F4	F2	F5	F3	F1
L10	M2	M1	M4	M3	M5	F2	F1	F4	F5	F3
L11	M2	M1	M4	M5	M3	F2	F5	F4	F3	F1
L12	M2	M4	M1	M5	M3	F4	F3	F5	F2	F1
L13	M2	M1	M4	M3	M5	F4	F1	F2	F3	F5

	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
Rank:	2	1	5	3	4	4	2	5	1	3
Tally:	31	24	53	35	52	48	32	51	21	43



Method: Quantifying Prosody

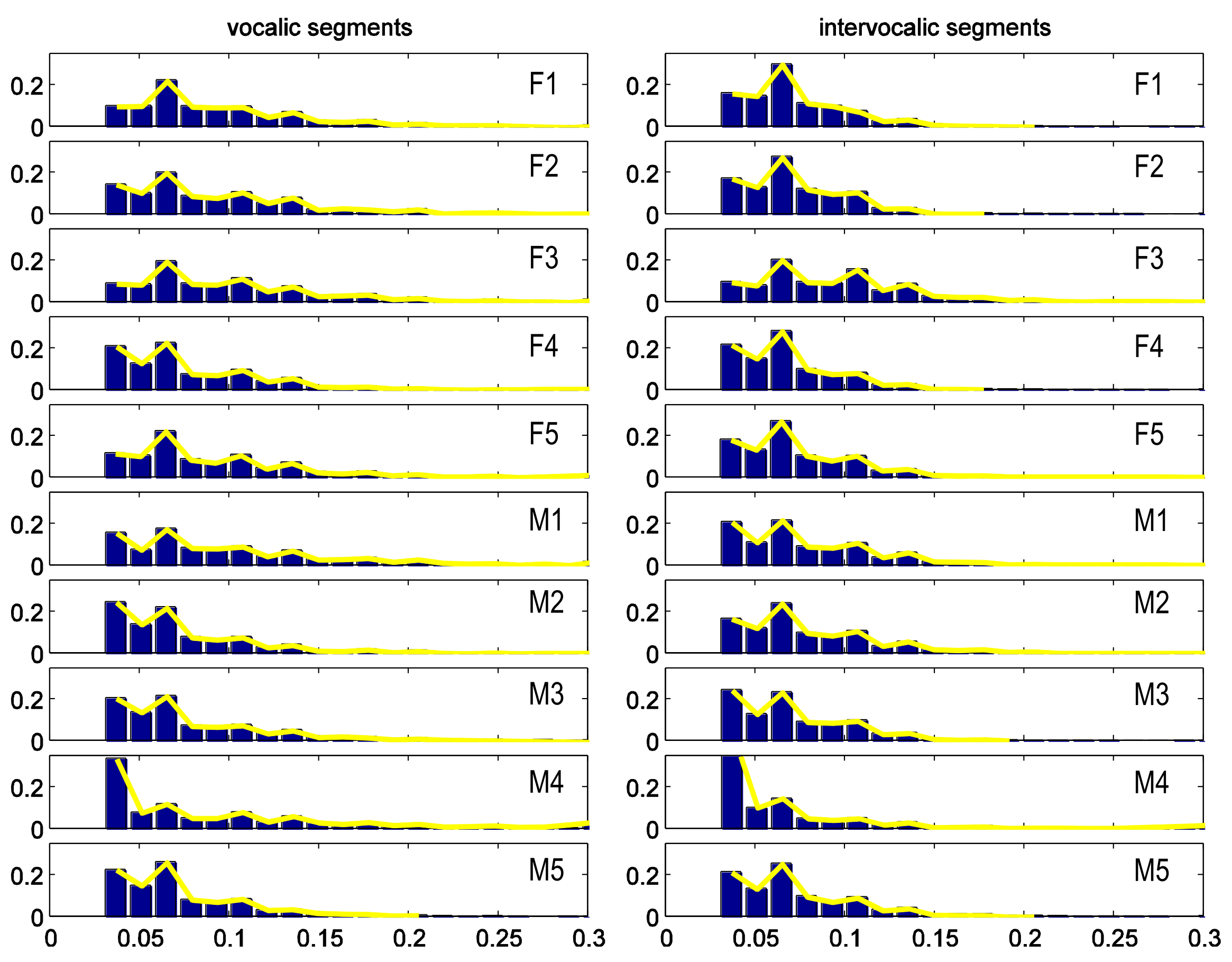
Audio samples preped for further analysis by forced-alignment phonetic transcription of each complete recording sampled in the listener survey.

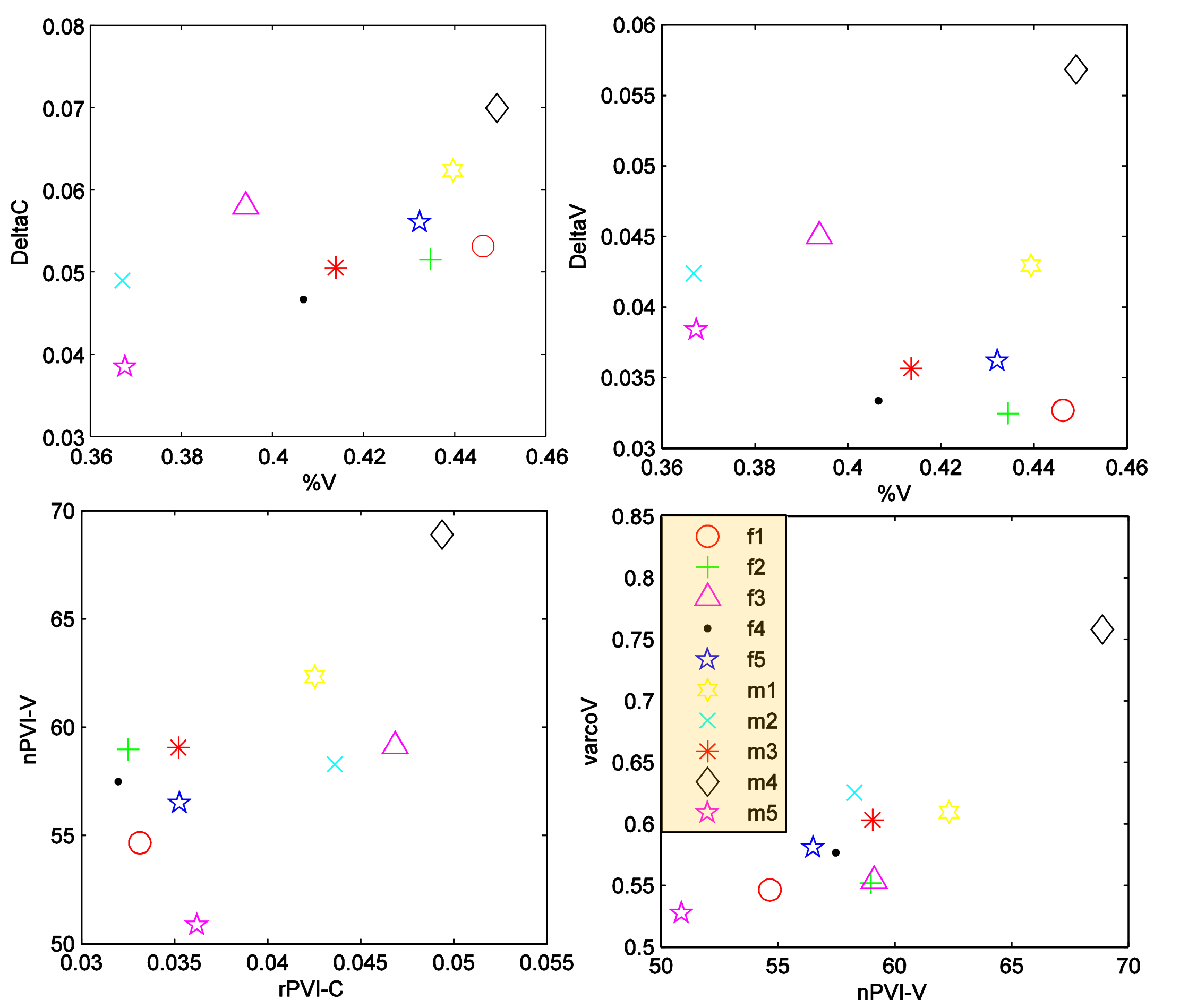
- companion texts sourced from LibriVox, Project Guttenberg [12,13]
- forced alignment using SailAlign: adaptive, iterative speech recognition & text alignment facilitating processing of audiobook-length speech recordings, and robust to transcription errors [14]
- transcriptions and interval timings generated at sentence-, word-, and phoneme-based levels of analysis

To compare the prosodic characteristics of each reader's speaking style, metrics were calculated for each text and reader including:

- percentage of vowels or vocalic intervals (%V)
- coefficient of variation of vocalic intervals (VarCoV)
- coefficient of variation of intervocalic intervals (VarCoC)
- normalized pair-wise variability index (nPVI)

Results: Reader Prosody





Conclusions

- listener responses to read speech are varied and complex, reflecting individual preferences which cannot always be identified or quantified
- nevertheless, some readers are consistently preferred amongst a population of native English speaking listeners; other reading voices are consistently identified as less felicitous
- standard metrics for quantifying prosodic properties of speech failed to robustly characterize readers as more or less felicitous, consistent with the intuitions of auditors
- more work is required to develop metrics capable of capturing properties of read speech which listeners are sensitive to

Future Directions

- broader survey of reading styles:
  - more listeners
  - more samples within and across literary genres
- control for specific prosodic and extra-prosodic factors through selection or manipulation of reading voices
- cross-language listener comparisons: native speakers of syllable-timed vs. foot-timed languages
- more sophisticated metrics capable of capturing super-segmental features of speech in multiple dimensions

References

[1] G. Laan (1997). *The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style*. Speech Communication 22(1):43-65

[2] R. Remez, P. Rubin, L. Nygaard (1986*). On spontaneous speech and fluently-spoken text: Production differences and perceptual distinctions*. JASA 79(S1): 26

[3] F. van Beinum (1991). *Spectro-temporal reduction and expansion in spontaneous speech and read text: Focus words versus non-focus words*. Proc. Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication: 36.1-36.5

[4] E. Blaauw (1995). *On the perceptual classification of spontaneous and read speech*. Doctoral dissertation, Utrecht University.

[5] M. Eskenazi (1993). *Trends in speaking styles research*. Proc. Eurospeech '93: 501-509

[6] P. Howell, K. Kadi-Hanifi (1991). Comparison of prosodic properties between read and spontaneous speech material. Speech Communication 10(2):163-169

[7] G. Fant, A. Kruckenberg, L. Nord (1991). *Some observations on tempo and speaking style in Swedish text reading*. Proc. Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication: 36.1-36.5

[8] D. Stojanovic (2009). Issues in the quantitative approach to speech rhythm comparisons. Working Papers in Linguistics 40(9):

[9] E. Grabe, E. L. Low (2003). *Durational variability in speech and the rhythm class hypothesis*. Papers in Laboratory Phonology (7): 515-546

[10] F.Ramus, M. Nespor, J. Mehler (1999). *Correlates of linguistic rhythm in the speech signal*. Cognition (73): 265-292

[11] J. Yuan, M. Liberman (2008). *Vowel acoustic space in continuous speech: An example of using audio books for research*. Cat-Cod

[12] J. London (1906). *White Fang.* Source: http://www.gutenberg.org/ebooks/23976

[13] J. London (1903). *The Call of the Wild.* Source: http://librivox.org/call-of-the-wild-by-jack-london/

[14] A. Katsamanis, M. Black, P. Georgiou, L. Goldstein, S. Narayanan (2011). *SailAlign: Robust long speech-text alignment*. Proc. New Tools and Methods for VLSPR, UPenn:

Acknowledgements

Research supported by NIH Grant R01 DC007124-01