



Characteristics of Text-to-Speech and Other Corpora

Erica Cooper, Emily Li, Julia Hirschberg

Columbia University, New York, USA

Research Questions

TTS data collection - typically:

- ▶ Professional voice talents
- ▶ Instructed to speak as clearly and consistently as possible
- ▶ Low variance in voice quality, speaking style, pitch, volume, or tempo

Do TTS speakers exhibit these characteristics more than speakers for other genres?

Which of these characteristics are important for TTS?

Which other genres are most similar to TTS data?

Data and Tools

TTS Research Corpora: ARCTIC (English), SWARA (Romanian), IIIT-H (Indic)

Other Professional Speech: TUNDRA (audiobooks), BURNC (radio news), Turkish broadcast news

ASR Corpora: CALLHOME (conversational), MACROPHONE (read), BABEL (LRLs)

Features: Mean and standard deviation of f0 and energy; speaking rate; articulation; NHR, jitter, and shimmer.

Results

Comparison p-value

slt vs. clb (ARCTIC) 0.336

slt vs. CALLHOME 5.65E-245

clb vs. CALLHOME 1.01E-245

Sign tests: ARCTIC female speakers ('slt' and 'clb') vs. each other and vs. CALLHOME for SD of f0

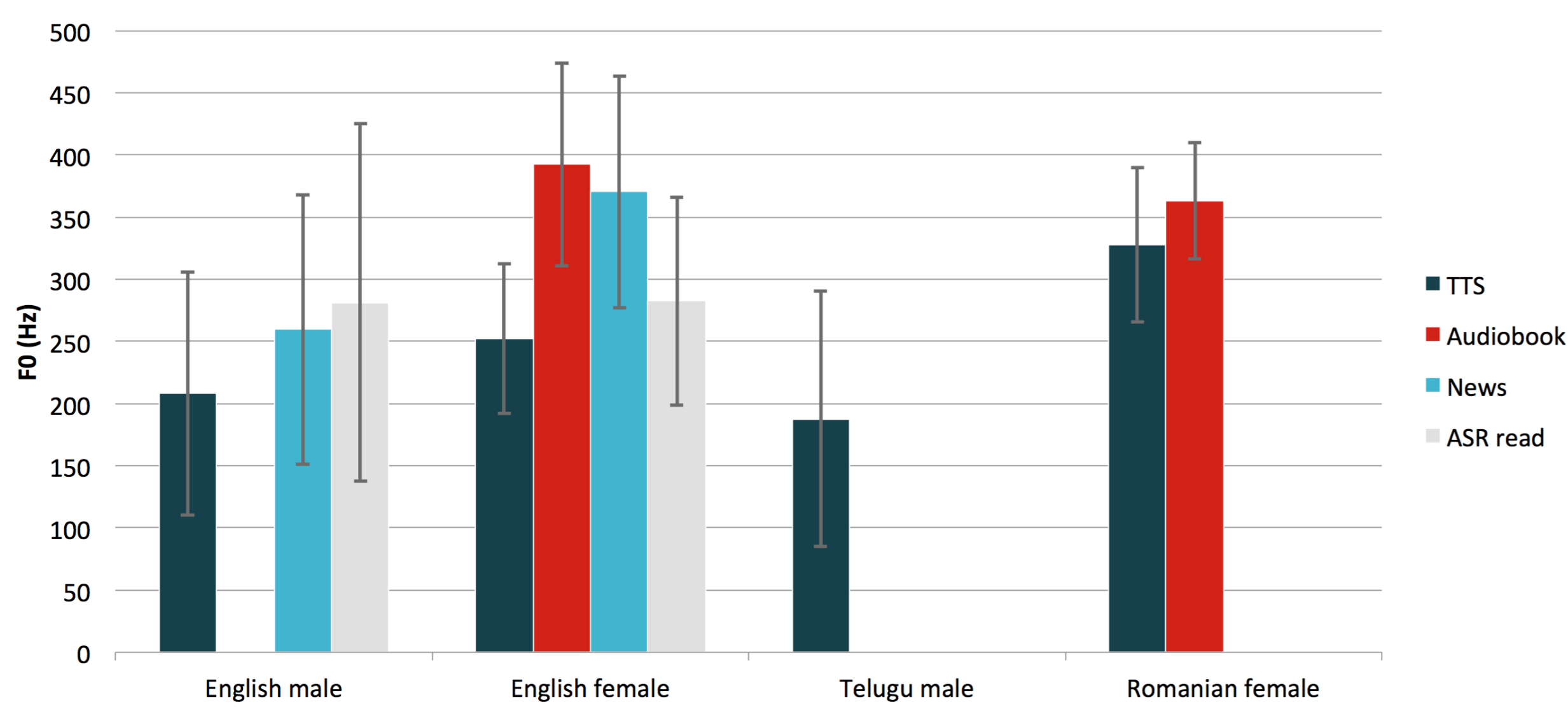
- ◆ Arctic male
- Arctic female
- ▲ SWARA male
- SWARA female
- IIIT_H male
- IIIT_H female

- ▲ Tundra male
- ◆ Tundra female
- Tundra male 1hr
- Tundra female 1hr

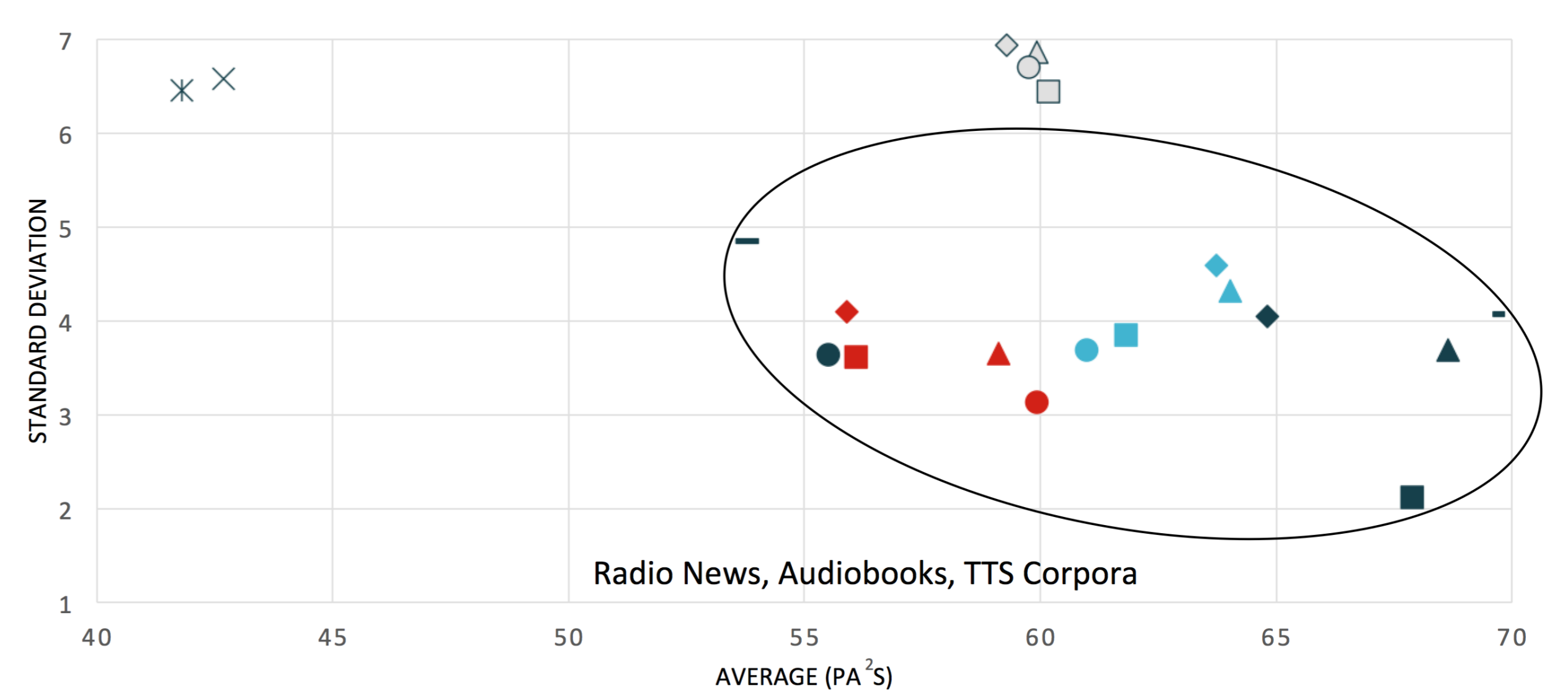
- BURNC male
- BURNC female
- ◆ TBN male
- ▲ TBN female

- × Macrophone male
- × Macrophone female
- △ Babel male
- Babel female
- ◇ Callhome male
- Callhome female

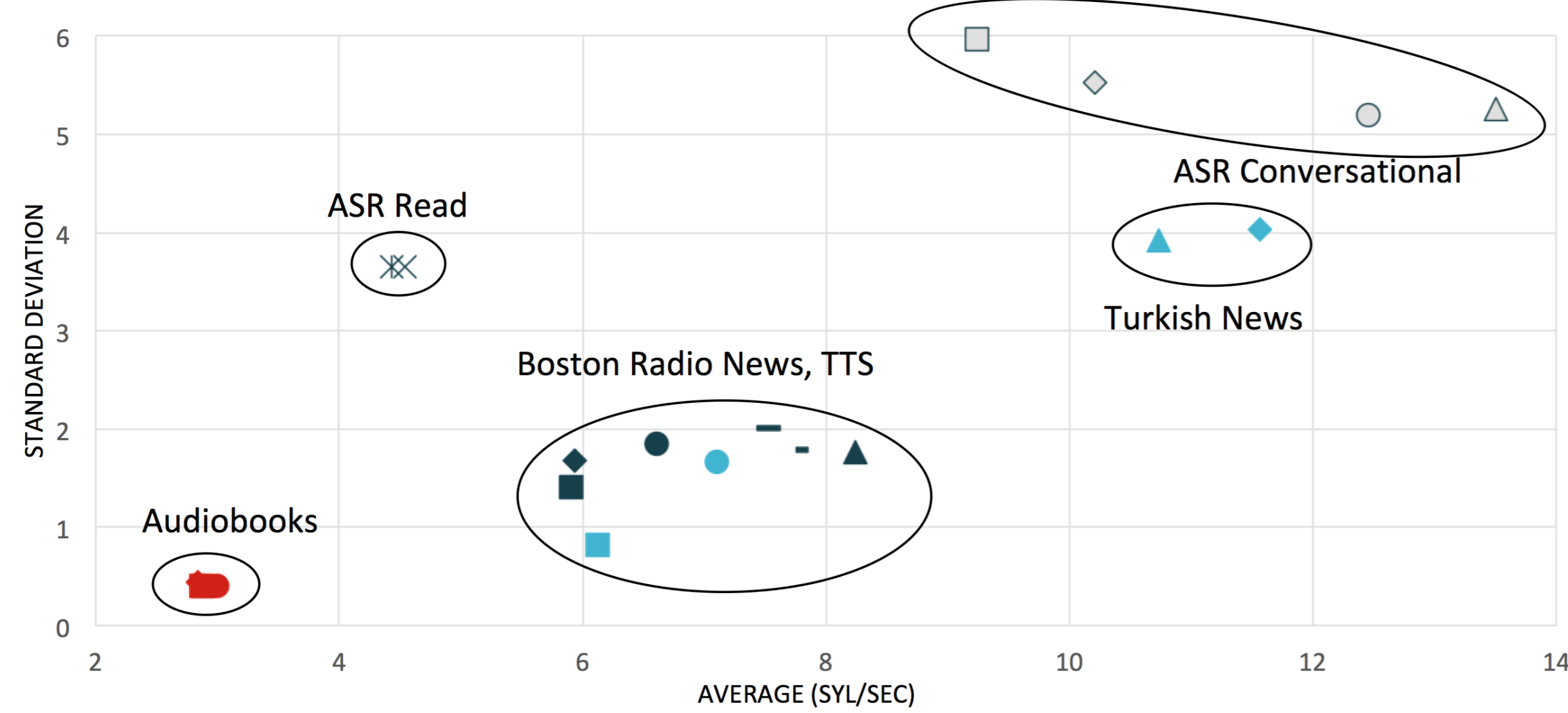
MEAN F0 FOR EACH GENRE BY LANGUAGE AND GENDER



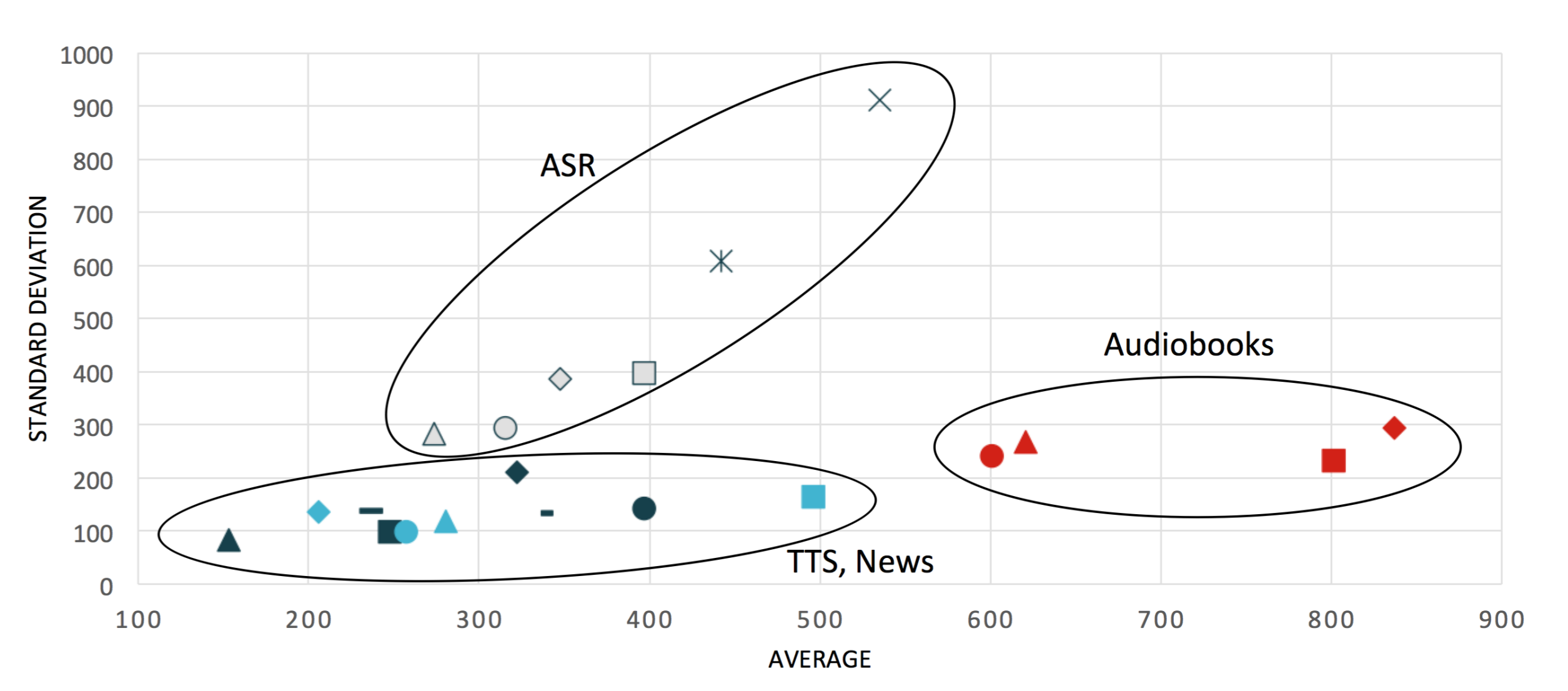
ENERGY AVERAGE VS. STANDARD DEVIATION



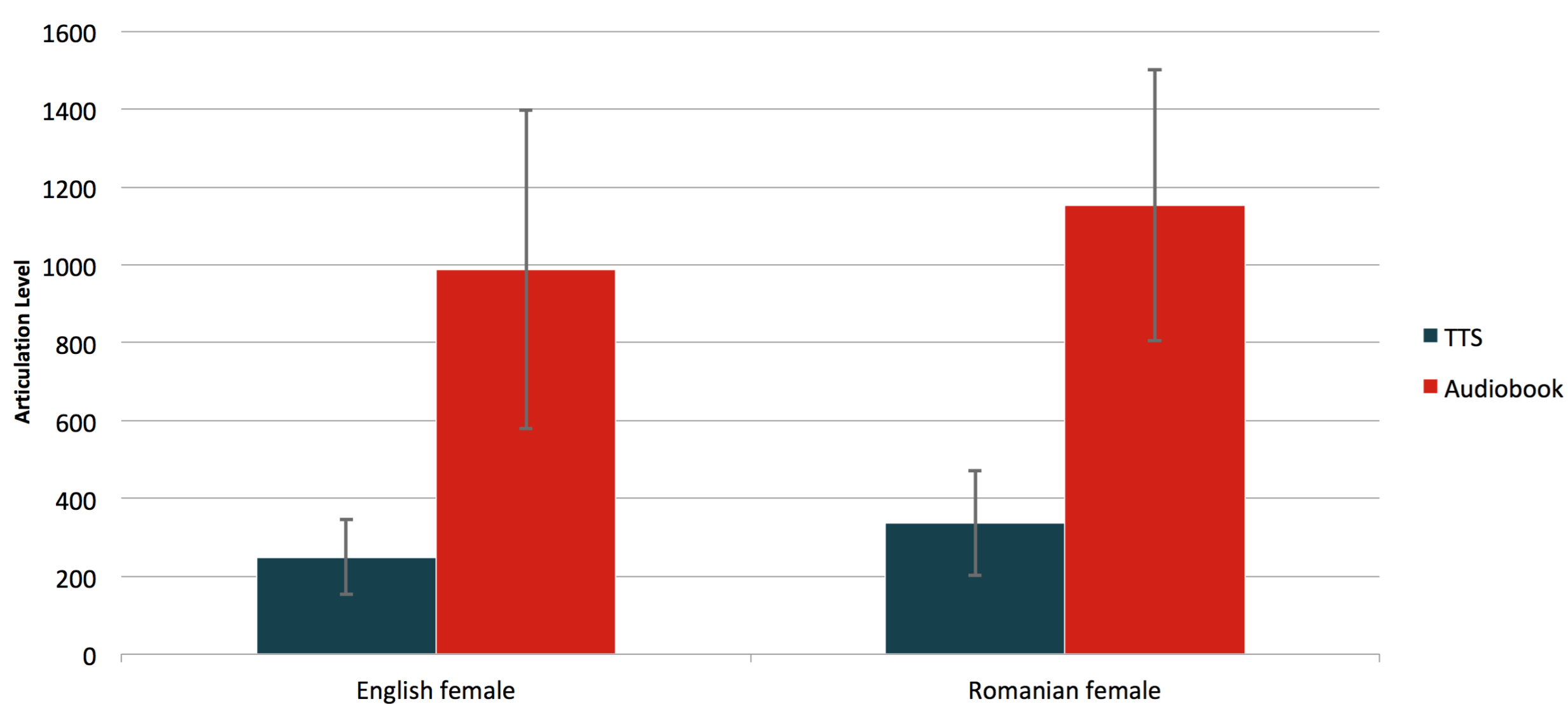
SPEAKING RATE AVERAGE VS. STANDARD DEVIATION



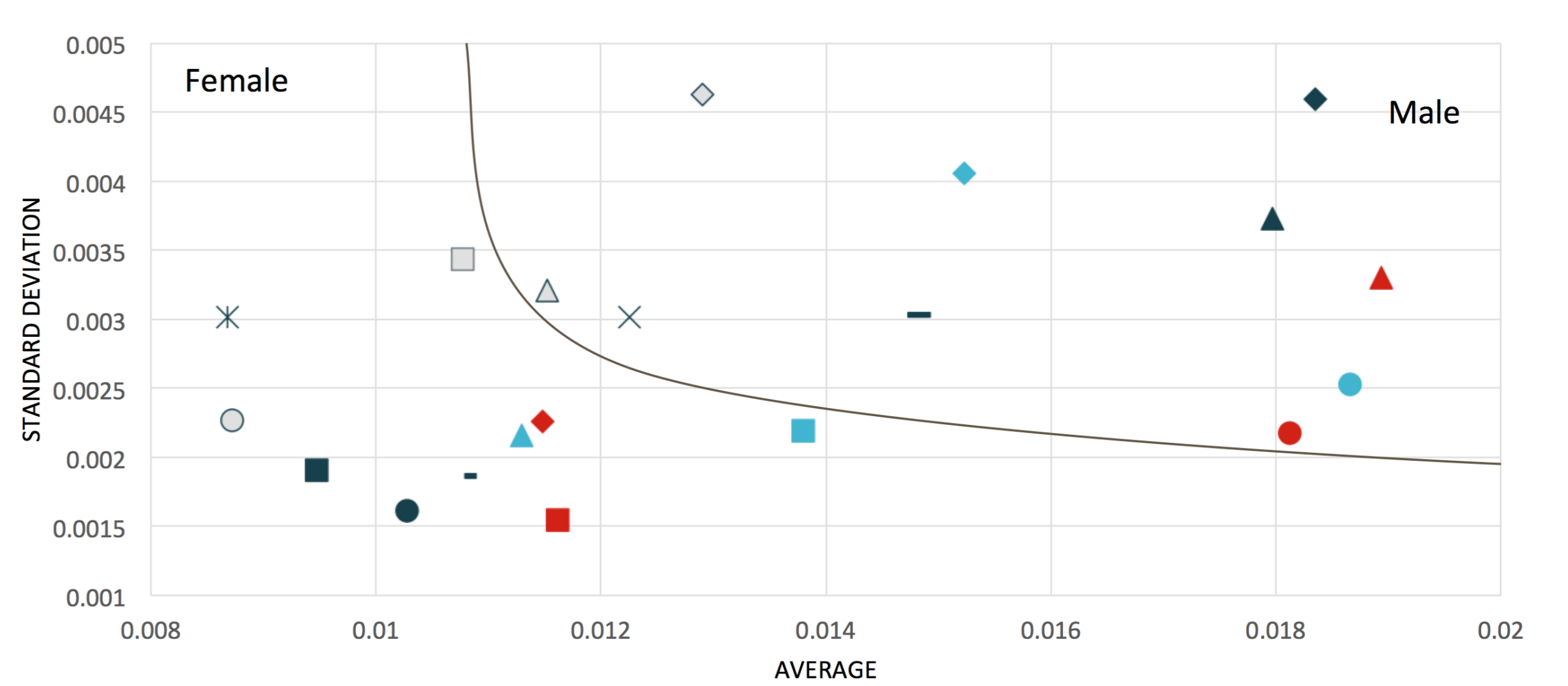
ARTICULATION AVERAGE VS. STANDARD DEVIATION



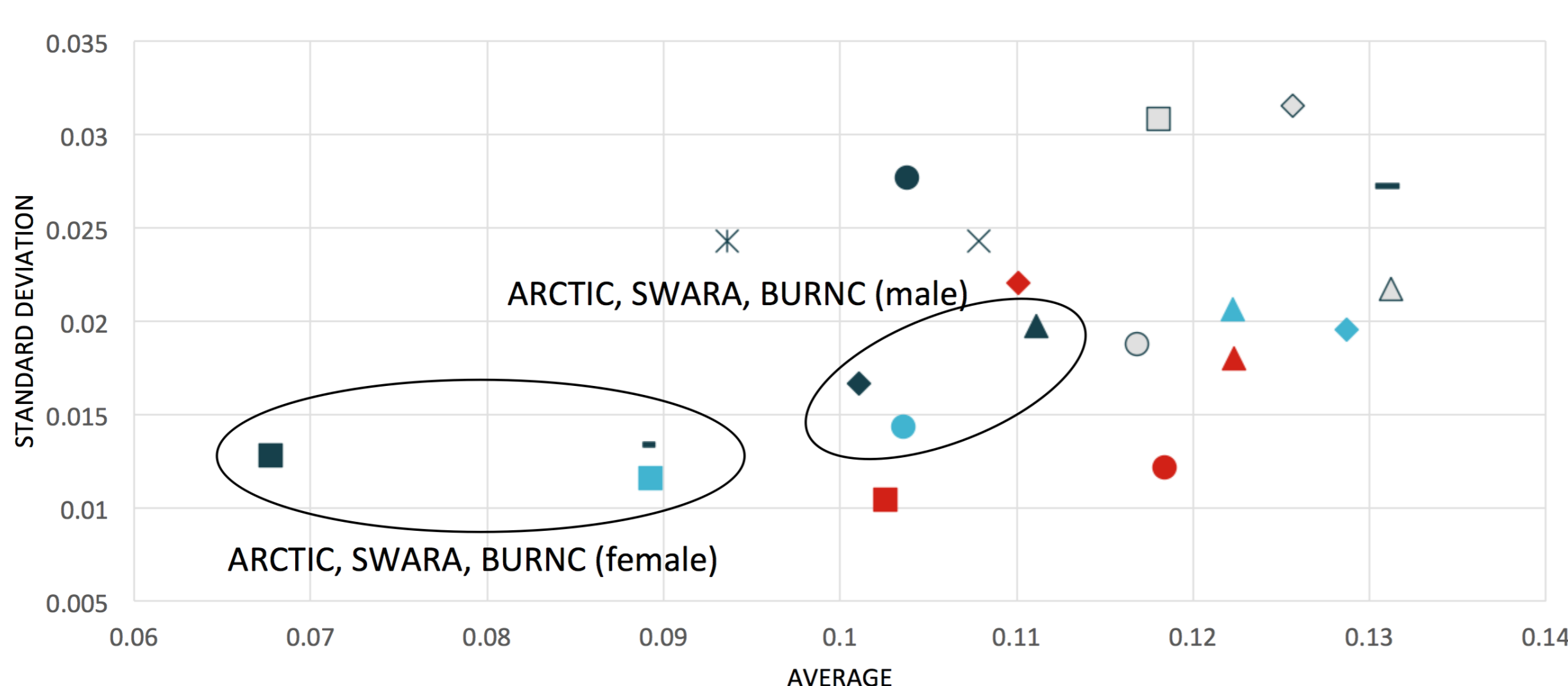
ARTICULATION FOR TTS AND AUDIOBOOK DATA



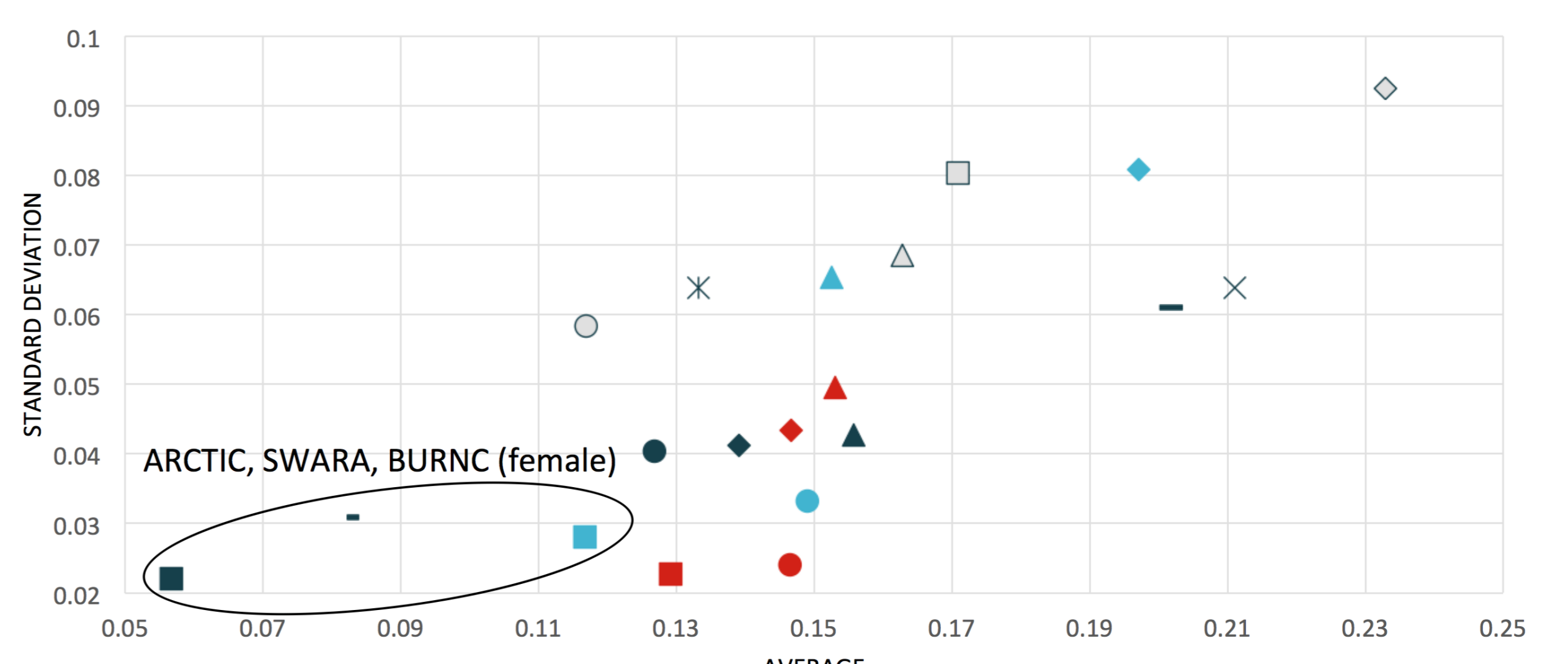
JITTER AVERAGE VS. STANDARD DEVIATION



SHIMMER AVERAGE VS. STANDARD DEVIATION



NHR AVERAGE VS. STANDARD DEVIATION



Acknowledgments

This work was supported by the National Science Foundation under Grants IIS 1548092 and 1717680.