Talker and language variation in the F0 of English, Mandarin and Mandarin-accented English

Lauren Ackerman, Lisa Hesterberg & Ann R. Bradlow

Northwestern University

MidPhon17
21 October 2011
Roadmap

- Background
- Methods
- Previous Work
- Analysis
- Discussion
Roadmap

- Background
- Methods
- Previous Work
- Analysis
- Discussion
Non-native speakers show a lot of variation in their speech.

Where does this variation come from?

We know some sources of variation:

- **L1-transfer**
  - e.g. Best et al. 1988, Flege 1986
- **Proficiency**
  - e.g. Flege 1995
- **Physiology**
  - e.g. Stevens 1998

How do these affect the speech signal?
Language Variation

We examine:

- Global features (not phonological features)
  - Phonological inventories are a source of variation
  - But how do they vary? (e.g., energy in LTASS; Byrne 1994)

How much of the variation comes from L1?

- Some variation is from L1-transfer (i.e. language-specific)
- Some variation is physiological (i.e. talker-specific)
  - e.g., Talkers with low-pitched voices in L1 will have low-pitched voices in all languages
- Which sources account for which observed variation?
Methodology: Materials

Recordings from ALLSSTAR

- **ALLSSTAR**: Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings
- Corpus designed for cross-linguistic and within-talker comparisons

Scripted recordings:
- The North Wind and the Sun (NWS)
  - a short read passage

Spontaneous recordings:
- Free response with prompt questions (QNA)
  - 5 minutes of spontaneous speech
Methodology: Subjects

- Foreign graduate students from the International Summer Institute (ISI) 2010
  - Speech stimuli controlled across languages
  - Recorded in L1 and L2

- American undergraduate students from Intro Linguistics classes
  - Monolingual English speakers
  - Recorded in L1

- We took recordings of Mandarin-speaking and English-speaking males only:
  - Largest *homogenous* group
  - Homogeny important for controlling physiology
Methodology: Subjects

Mandarin talkers
- speaking L1 Mandarin: **MM**; N=11
- speaking L2 English: **ME**; (same talkers)
- ages 22 to 26, mean = 23.5 years
- length of residency = about 1 month
- VERSANT test scores (spoken proficiency)
  - 50 to 59, mean 55.9 (scale of 20-80)

English talkers
- speaking L1 English: **EE**; N=8
- reported as monolingual
- ages 18 to 26, mean = 20 years
Roadmap

- Background
- Methods
- **Previous Work**
- Analysis
- Discussion
Previous Analyses

We examined LTASS:

- long-term average speech spectra
- overall measure of energy in the speech signal
- energy averaged over whole recording in 50 Hz bins

We found:

- MM has lower energy than EE; ME is between the two
- QNA has higher energy than NWS; stronger effect in MM
- Higher proficiency is correlated with higher energy (e.g. more EE-like)

Lower range (< 200 Hz) has multiple sources of variation:

- F0, vowel formants, nasals, etc.
Previous Studies

Loveday 1981:
- English and Japanese, plus Japanese-accented English
- Differences found in speech style across languages (within and across subjects)
- But: small sample sizes ($N_{\text{male}} = 3$, $N_{\text{female}} = 2$)

Eady 1982:
- Compared L1 Mandarin and L1 English:
  - mean F0
  - standard deviation of F0
  - speech rate
  - pitch fluctuation
  - Mandarin and English differ in mean F0, speech rate, etc
- But: no Mandarin-accented (L2) English
  - only scripted speech
Roadmap

- Background
- Methods
- Previous Work
- **Analysis**
- Discussion
We use full-to-subset linear mixed effects regression models compared using ANOVAs

Mean F0:
- The average frequency the talker uses in a speech sample
- This gives us an impression of overall pitch used in a talker’s speech

Standard Deviation of F0:
- Our pitch extraction algorithm has difficulty with min/max, so StDev is a stand-in for range

Proficiency:
- Proficiency in our sample does not vary greatly, but it may shed light on any acoustic variation
Mean F0: By Language

Mean F0 Across Languages

* (*

Measures F0 for different languages:
- EE: Red
- ME: Purple
- MM: Blue

Significance:
- n.s.: Not significant

Talker and language variation in F0
Mean F0: By Task

Mean F0 Across Tasks

NWS

QNA

Talker and language variation in F0

Lauren Ackerman, Lisa Hesterberg & Ann R. Bradlow
Mean F0: Interaction

Mean F0 Across Languages and Tasks

- EE NWS
- EE QNA
- ME NWS
- ME QNA
- MM NWS
- MM QNA

Talker and language variation in F0
Mean F0: Interaction

Mean F0 Across Languages and Tasks

![Box plot showing mean F0 across languages and tasks.](image-url)
Mean F0: Interaction

Mean F0 Across Languages and Tasks

- EE NWS
- EE QNA
- ME NWS
- ME QNA
- MM NWS
- MM QNA

Fundamental Frequency (Hz)

Task by Language
Standard Deviation of F0

Standard Deviation of F0 Across Languages and Tasks

Task by Language:
- EE NWS
- EE QNA
- ME NWS
- ME QNA
- MM NWS
- MM QNA

Fundamental Frequency (Hz)

Lauren Ackerman, Lisa Hesterberg & Ann R. Bradlow
Talker and language variation in F0
Within-talker Correlations

- ME & MM correlate ($r[8]=0.63$, $p < 0.05$)
- ME & VERSANT score correlate ($r[8]=-0.80$, $p < 0.05$)
- MM & VERSANT score do not correlate ($r[8]=-0.33$, $p > 0.1$)
Within-talker Correlations

- ME & MM correlate \((r[8]=0.63, p < 0.05)\)
- ME & VERSANT score correlate \((r[8]=-0.80, p < 0.05)\)

Lauren Ackerman, Lisa Hesterberg & Ann R. Bradlow

Talker and language variation in F0
Within-talker Correlations

- ME & MM correlate ($r[8]=0.63$, $p < 0.05$)
- ME & VERSANT score correlate ($r[8]=-0.80$, $p < 0.05$)
- So does the test simply prefer lower-pitched voices?
Within-talker Correlations

- ME & MM correlate \( r[8]=0.63, p < 0.05 \)
- ME & VERSANT score correlate \( r[8]=-0.80, p < 0.05 \)
- So does the test simply prefer lower-pitched voices?
  - MM & VERSANT score do not correlate \( r[8]=-0.33, p > 0.1 \)
Roadmap

- Background
- Methods
- Previous Work
- Analysis
- Discussion
Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely
- LTASS: QNA has higher energy than NWS; stronger in MM
- F0: NWS has higher F0 than QNA; strongest in MM
- LTASS: Higher proficiency correlated with higher energy (i.e., more EE-like)
- F0: Higher proficiency correlated with lower pitch (i.e., more EE-like)

F0 shows cross-linguistic variation, matching LTASS findings.

ME is more similar to EE than it is to MM.

L1 Mandarin speech shows the most robust task differences.
Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely
Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely
- LTASS: QNA has higher energy than NWS; stronger in MM
- LTASS: Higher proficiency correlated with higher energy (i.e., more EE-like)
- F0: Higher proficiency correlated with lower pitch (i.e., more EE-like)

F0 shows cross-linguistic variation, matching LTASS findings.

ME is more similar to EE than it is to MM.

L1 Mandarin speech shows the most robust task differences.
F0 and LTASS

Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely

- LTASS: QNA has higher energy than NWS; stronger in MM
- F0: NWS has higher F0 than QNA; strongest in MM
F0 and LTASS

Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely

- LTASS: QNA has higher energy than NWS; stronger in MM
- F0: NWS has higher F0 than QNA; strongest in MM

- LTASS: Higher proficiency correlated with higher energy (i.e., more EE-like)
F0 and LTASS

Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely
- LTASS: QNA has higher energy than NWS; stronger in MM
- F0: NWS has higher F0 than QNA; strongest in MM
- LTASS: Higher proficiency correlated with higher energy (i.e., more EE-like)
- F0: Higher proficiency correlated with lower pitch (i.e., more EE-like)
Combining results from LTASS and F0:

- LTASS: MM has lower energy than EE; ME is in between
- F0: MM has higher F0, ME matches EE closely

- LTASS: QNA has higher energy than NWS; stronger in MM
- F0: NWS has higher F0 than QNA; strongest in MM

- LTASS: Higher proficiency correlated with higher energy (i.e., more EE-like)
- F0: Higher proficiency correlated with lower pitch (i.e., more EE-like)

F0 shows cross-linguistic variation, matching LTASS findings.
Combining results from LTASS and F0:

- **LTASS**: MM has lower energy than EE; ME is in between
- **F0**: MM has higher F0, ME matches EE closely
- **LTASS**: QNA has higher energy than NWS; stronger in MM
- **F0**: NWS has higher F0 than QNA; strongest in MM
- **LTASS**: Higher proficiency correlated with higher energy (i.e., more EE-like)
- **F0**: Higher proficiency correlated with lower pitch (i.e., more EE-like)

1. F0 shows cross-linguistic variation, matching LTASS findings
2. ME is more similar to EE than it is to MM
Combining results from LTASS and F0:

- **LTASS**: MM has lower energy than EE; ME is in between
- **F0**: MM has higher F0, ME matches EE closely

- **LTASS**: QNA has higher energy than NWS; stronger in MM
- **F0**: NWS has higher F0 than QNA; strongest in MM

- **LTASS**: Higher proficiency correlated with higher energy (i.e., more EE-like)
- **F0**: Higher proficiency correlated with lower pitch (i.e., more EE-like)

1. F0 shows cross-linguistic variation, matching LTASS findings
2. ME is more similar to EE than it is to MM
3. L1 Mandarin speech shows the most robust task differences
Thank you!

Thanks to:

NIH Grant # R01 DC005794
Bradlow Lab, Kelsey Mok, Chun Liang Chan
ISI 2010

References:


Average LTASS Energy By Task and Language

Energy (dB)

200-1000 Hz  1000-3000 Hz  3000-8000 Hz

NWS  QNA