# Elucidate the elusive concept of ‘pleasant voice’ for interpreters: a case study

**Caiwen Wang**
Translation and Interpreting Studies
University of Westminster

**Mark Huckvale**
Speech, Hearing and Phonetic Sciences
University College London

# What is a ‘Pleasant Voice’ for interpreters?

No clear definition in interpreting literature

# Literature review

No definition provided at the very beginning:

‘Pleasant voice’ stems from Bühler (1986), the first empirical study on criteria for assessing interpreting quality, followed by a series of studies along the same/similar line, e.g. Kurz (1993), Moser (1995), Collados Aís (1998), Chiaro and Nocella (2004), Zhang and Ke (2008), Pöchhacker and Zwischenberger (2010).

These ‘user expectation surveys suffer from methodological limitations, as criteria like completeness, correct grammar or pleasant VOICE may mean different things to different respondents.’ (Pöchhacker 2015: 431)

# Literature review

Definitions of ‘pleasant voice’ if provided:

**Either**:

in terms of whether or not it has unpleasant features of voice, e.g., Schjoldager (1996): squeaky, hoarse, weepy, unconvincing;

Iglesias Fernández (2007): high pitch & nasal timbre

pleasant voice = voice characteristics

# Literature review

**Or:**

‘Voice’ is confounded with delivery, either with no effort to distinguish, e.g. Gile (1990) where ‘quality of voice’ and ‘delivery’ are either considered together or hard to tell apart, e.g. Iglesias Fernández (2007) where respondents found it hard to separate delivery from voice.

pleasant voice = voice characteristics + delivery

# Literature review

**Or:**

‘Voice’ is confounded with delivery and speaker personality, e.g. , De Gregoris (2016) mentions articulation, hesitation, speed of speech, melodious/monotonous voice, sweet/aggressive voice, active/self-defeating personality, comprehensible voice, expressive voice, credible voice.

pleasant voice = voice characteristics + delivery + personality

# Teaching & Learning

- Imprecise definitions:
  - make it hard to explain to students what constitutes a pleasant voice
  - make it hard for students to know how to improve their performance

# Proposal

1. Present a framework for describing vocal performance
2. Investigate scientific methods for assessment
3. Initiate a study of listener preferences

# Descriptive Framework

**Linguistics:**
what is spoken

**Extralinguistics:**
who is speaking

**Paralinguistics:**
how is spoken

Following: John Laver (1994)

# Linguistic elements

- Verbal content
- Prosodic functions
  - Sentence functions (question/statement)
  - Word prominence (focus, emphasis)

The *interpretation* part of the performance

# Extralinguistic elements

- The speaker as a person
  - Identity
  - Personality
- The speaker as a member of a group
  - Accent (Regional & L2)

Not typically manipulated by the speaker

# Paralinguistic elements

- Speaking Style
  - Intelligibility
  - Formality
- Emotion
  - Arousal (quantity)
  - Valence (quality)

The focus for improving vocal performance

# Experimental study


- Computational paralinguistics
  - Collect corpus of labelled speech
  - Build model that predicts listener ratings from acoustic properties

Experimental basis for studying pleasant voice?

# Acoustic Measurement

**Phonetic properties**

- Intelligibility
  - Voice effectiveness
  - Articulatory quality
  - Speaking rate
  - Fluency
- Arousal
  - Pitch height & range
  - Voice quality

**Acoustic properties**

- Short-term temporal structure
  - regularity, noisiness
- Fundamental frequency
- Spectral shape
- Timing
- Temporal modulations

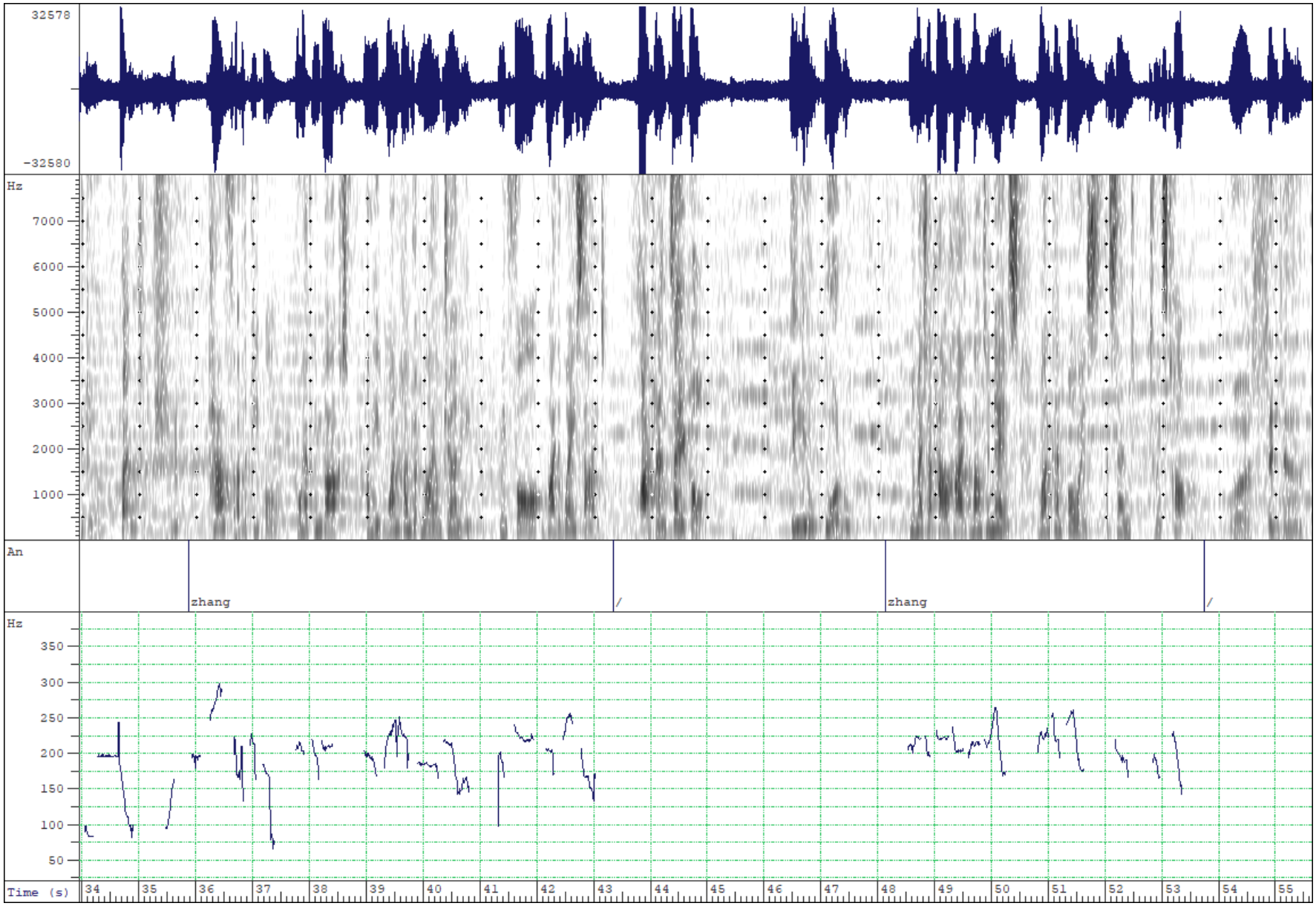
# Case Study: Zhang Lu


Zhang Lu, a chief interpreter with China's Ministry of Foreign Affairs


CREATED WITH
weVIDEO
FREE VERSION
和则两立 斗则俱伤

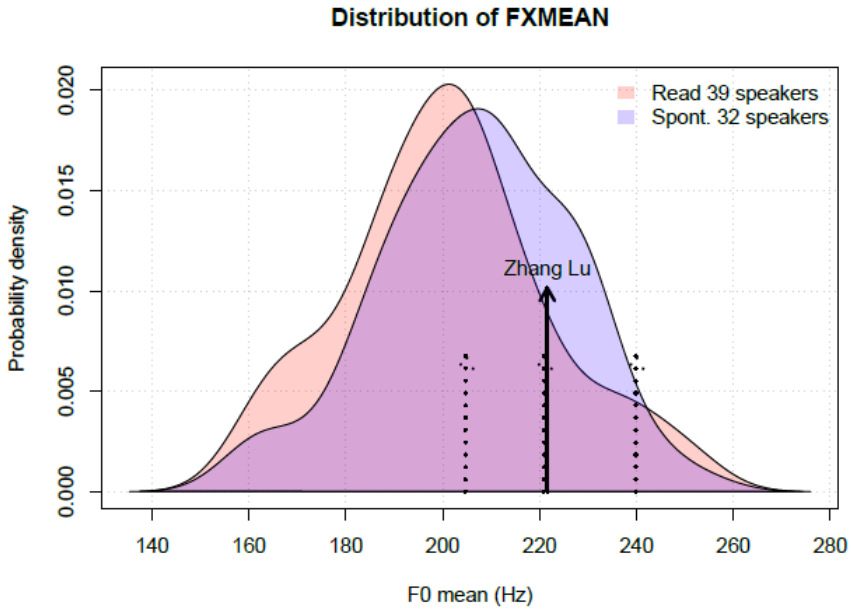
# Speech Materials

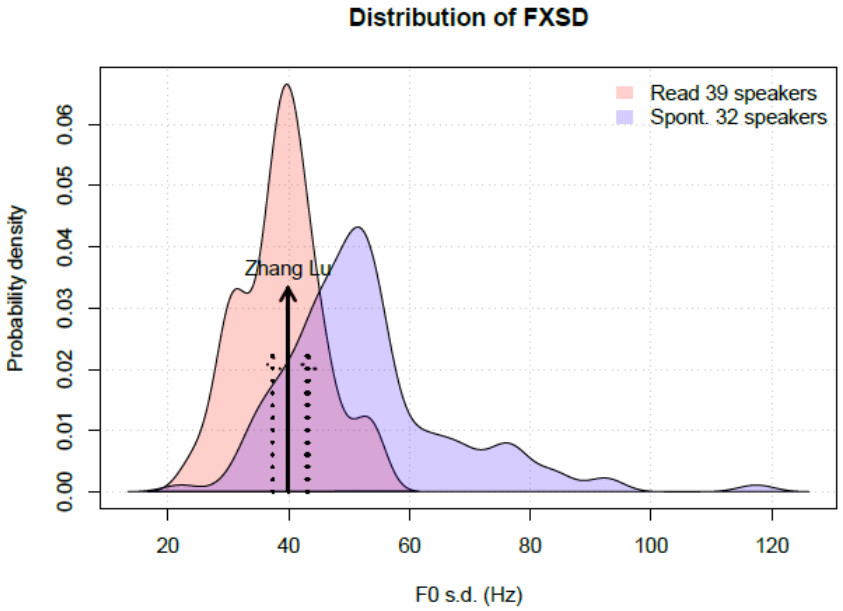
- Zhang Lu interpreting
  - 3 YouTube videos
- Corpus of read speech
  - WSJCAM0 corpus (39 female speakers)
- Corpus of spontaneous speech
  - HCRC Map Task corpus (32 female speakers)

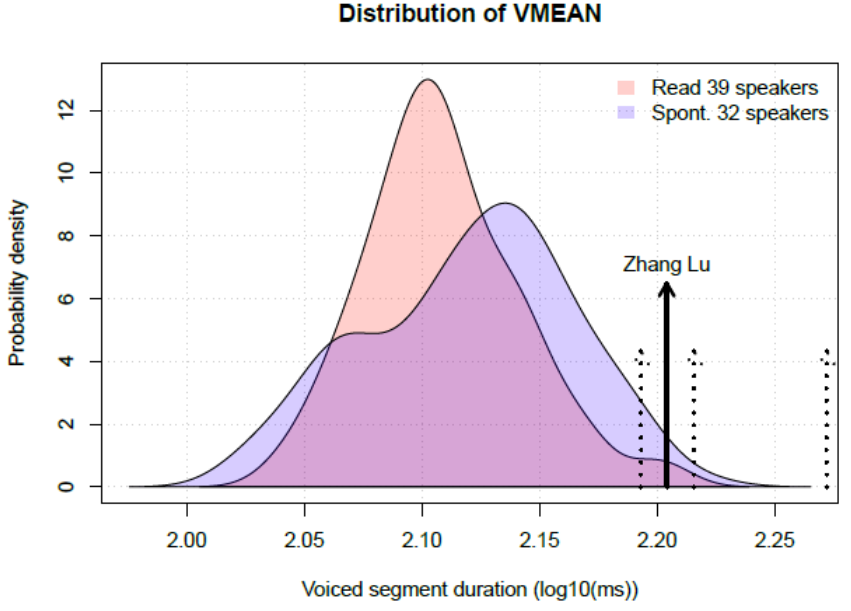

32578
-32580
Hz
7000
6000
5000
4000
3000
2000
1000
An
zhang
/
zhang
/
Hz
350
300
250
200
150
100
50
Time (s)
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55

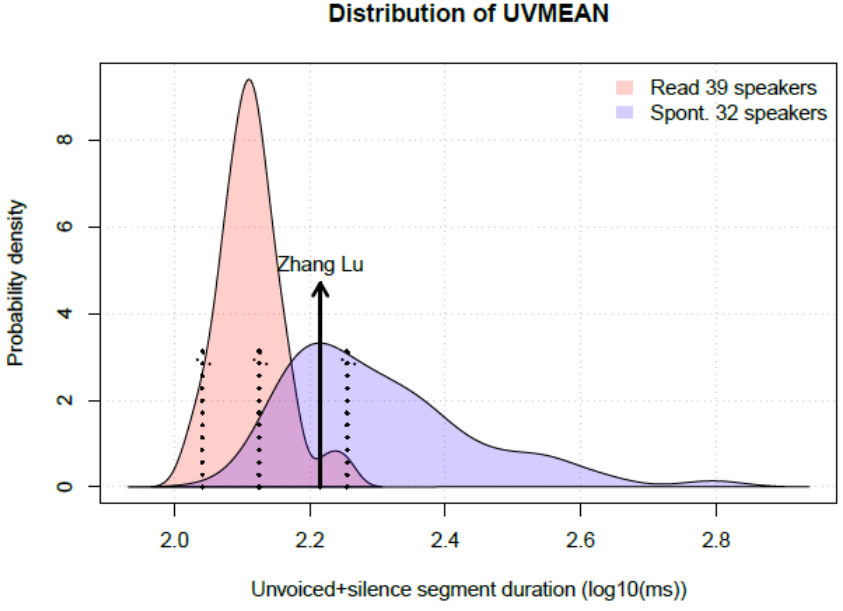
# Signal Analysis

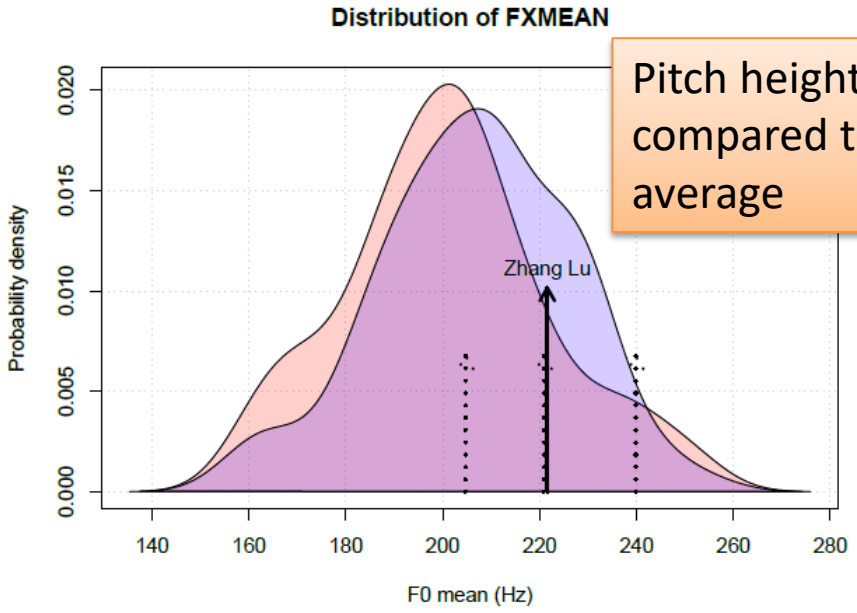
- Fundamental frequency analysis
  - F0 track by RAPT method
  - Voicing by autocorrelation
- Manual correction for recordings corrupted with music
- Collect F0 mean, F0 s.d., mean duration of voiced segments, mean duration of unvoiced segments and pauses
- Compare Zhang Lu's values with distribution of values by reference speakers


Distribution of FXMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0.000
0.005
0.010
0.015
0.020
140
160
180
200
220
240
260
280
F0 mean (Hz)

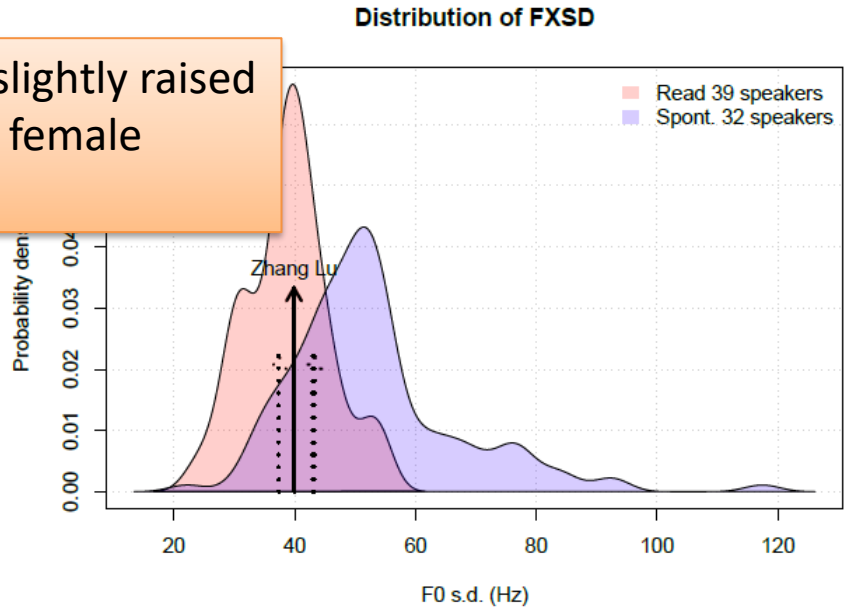

Distribution of FXSD
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
20
40
60
80
100
120
F0 s.d. (Hz)

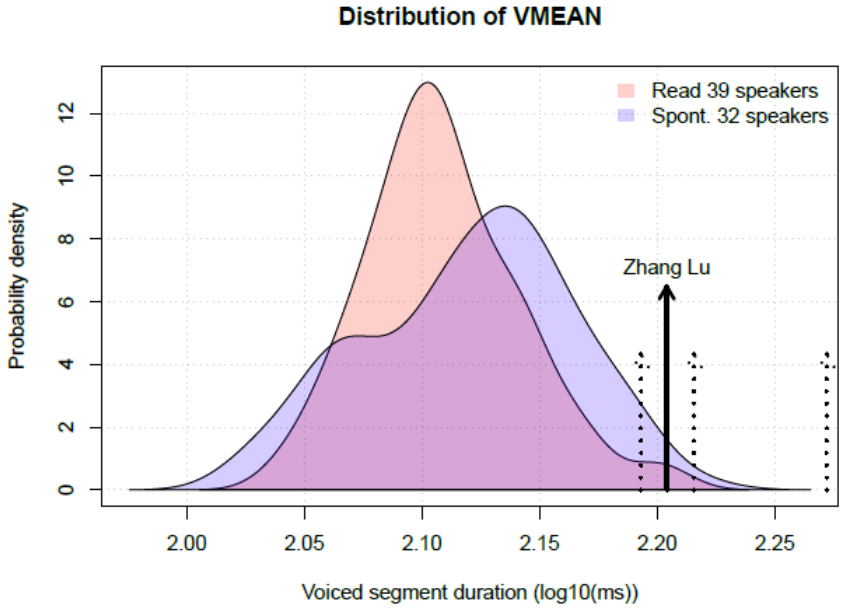

Distribution of VMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0
2
4
6
8
10
12
2.00
2.05
2.10
2.15
2.20
2.25
Voiced segment duration (log10(ms))

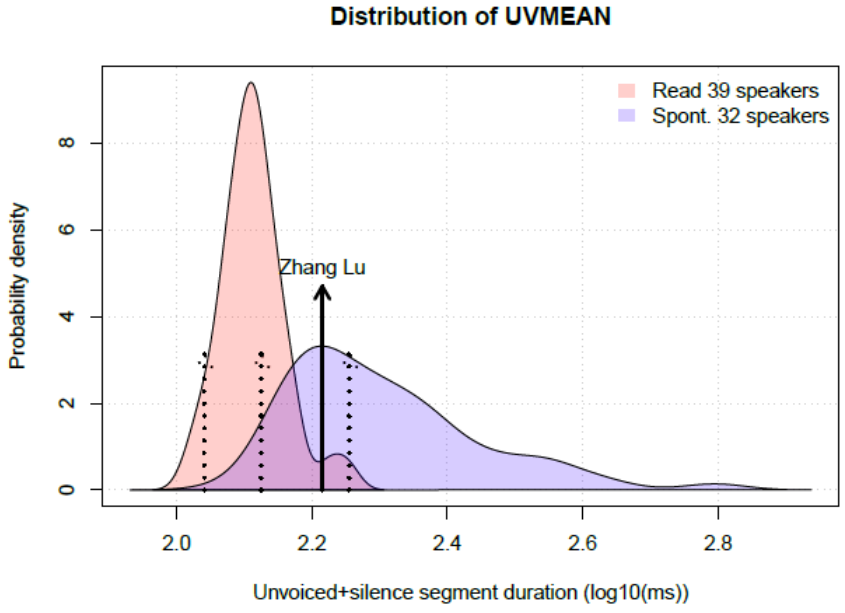

Distribution of UVMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0
2
4
6
8
2.0
2.2
2.4
2.6
2.8
Unvoiced+silence segment duration (log10(ms))

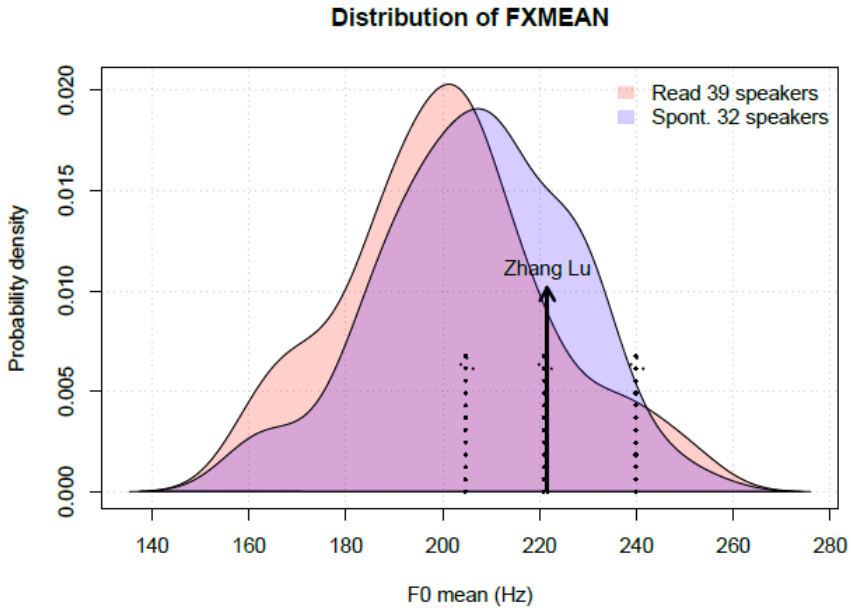


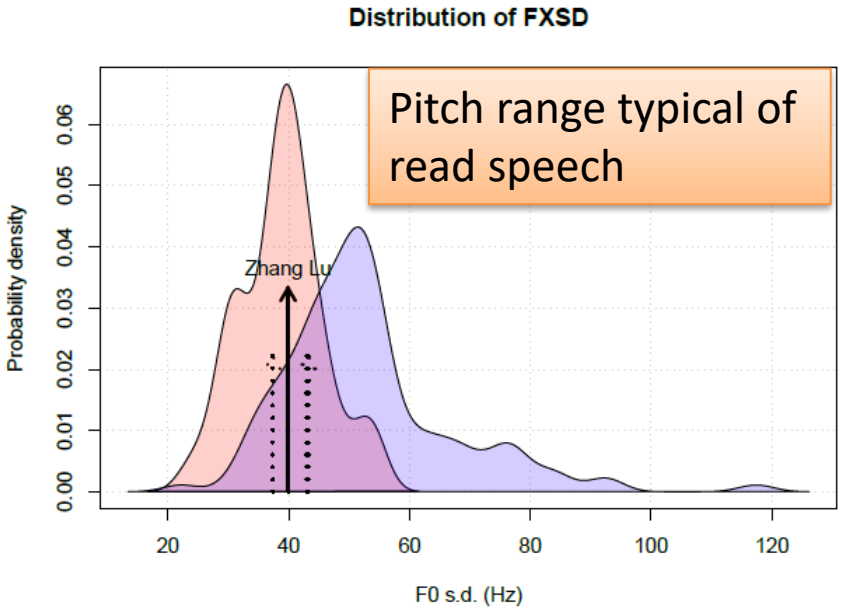

Pitch height slightly raised compared to female average

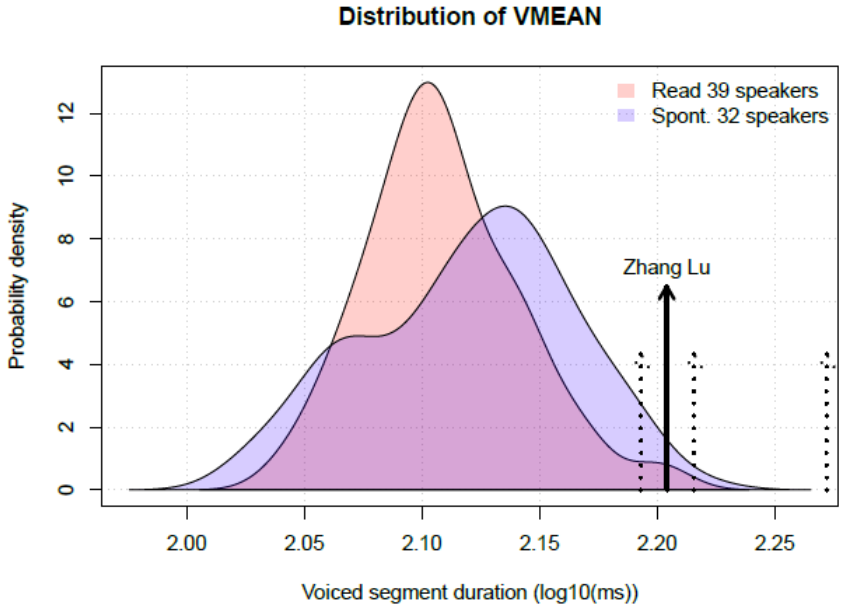


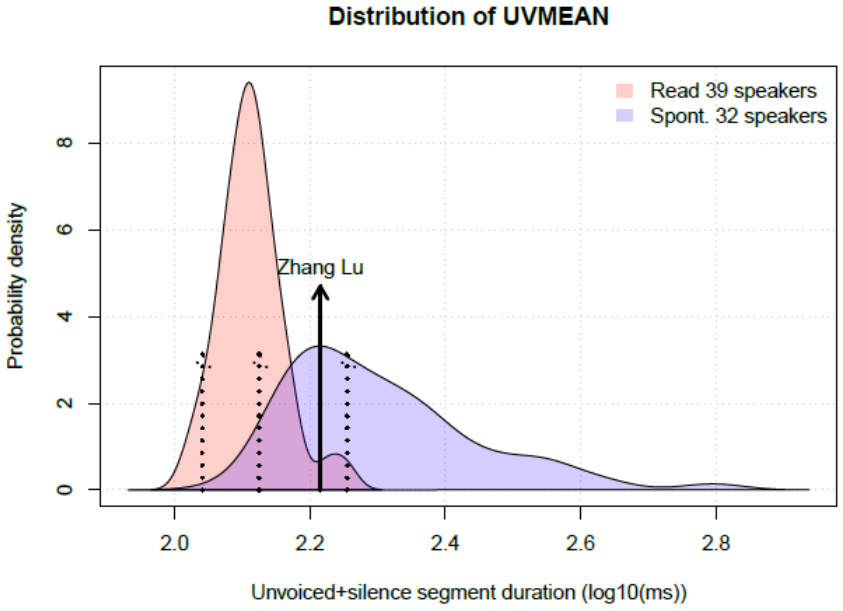



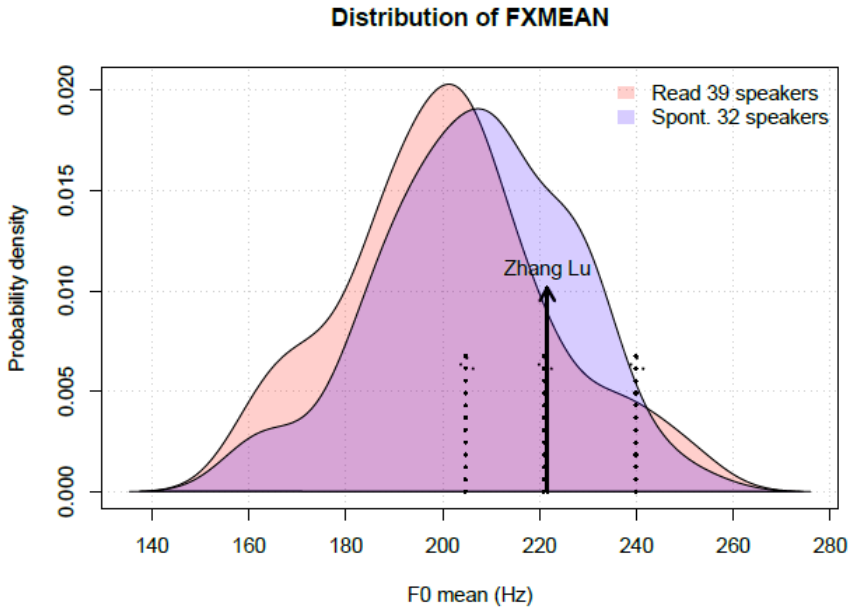



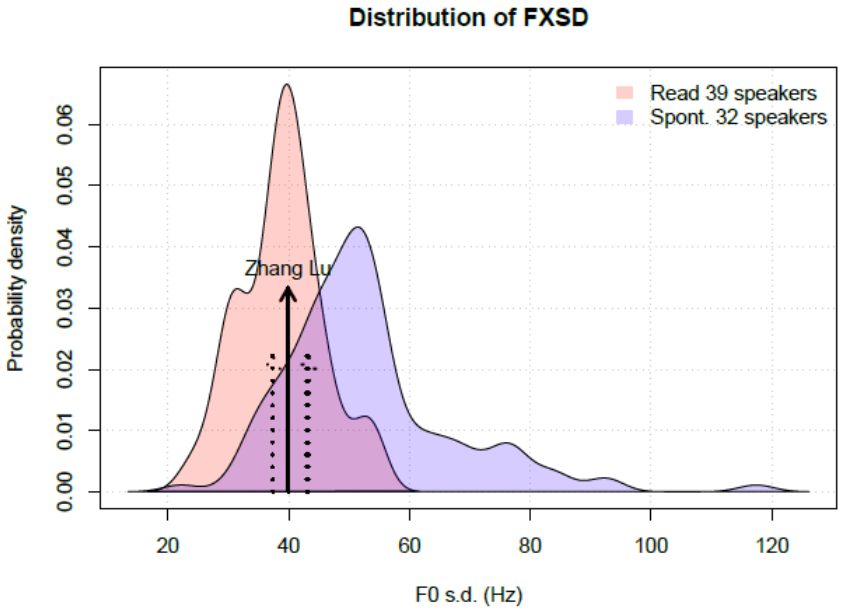

Distribution of FXMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0.000
0.005
0.010
0.015
0.020
140
160
180
200
220
240
260
280
F0 mean (Hz)

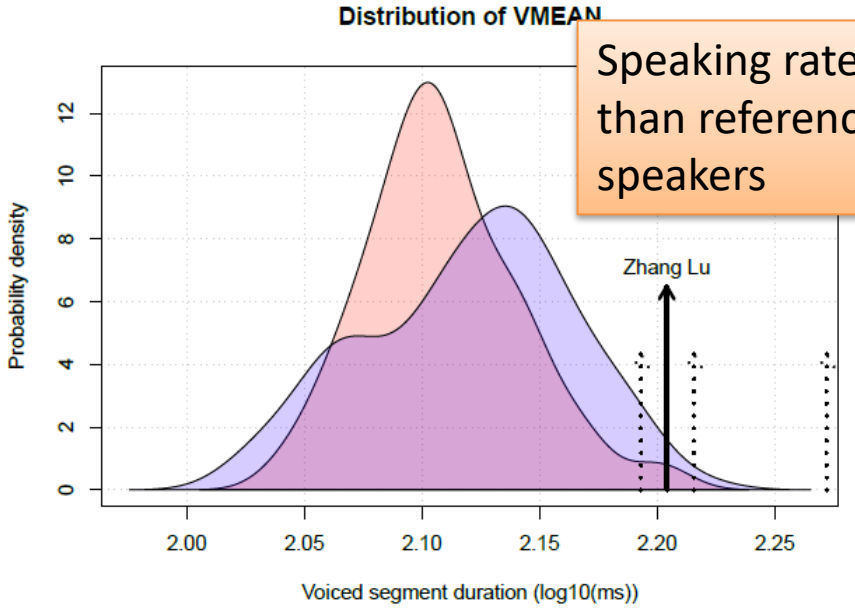

Distribution of FXSD
Pitch range typical of read speech
Zhang Lu
Probability density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
20
40
60
80
100
120
F0 s.d. (Hz)


Distribution of VMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0
2
4
6
8
10
12
2.00
2.05
2.10
2.15
2.20
2.25
Voiced segment duration (log10(ms))

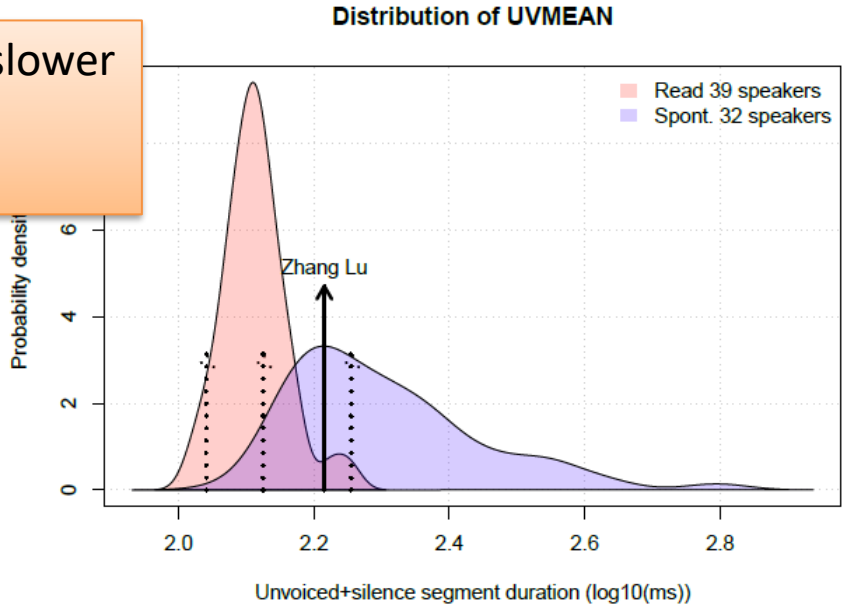

Distribution of UVMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0
2
4
6
8
2.0
2.2
2.4
2.6
2.8
Unvoiced+silence segment duration (log10(ms))

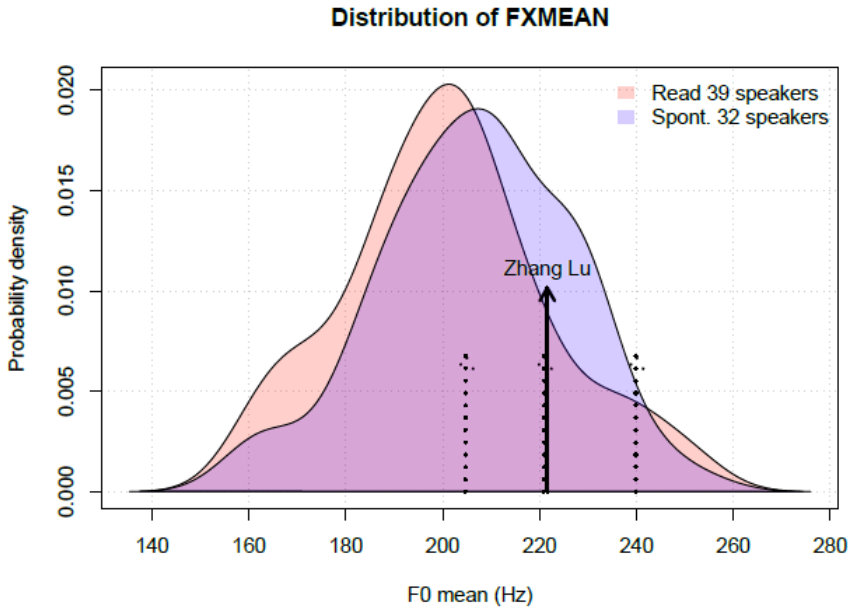


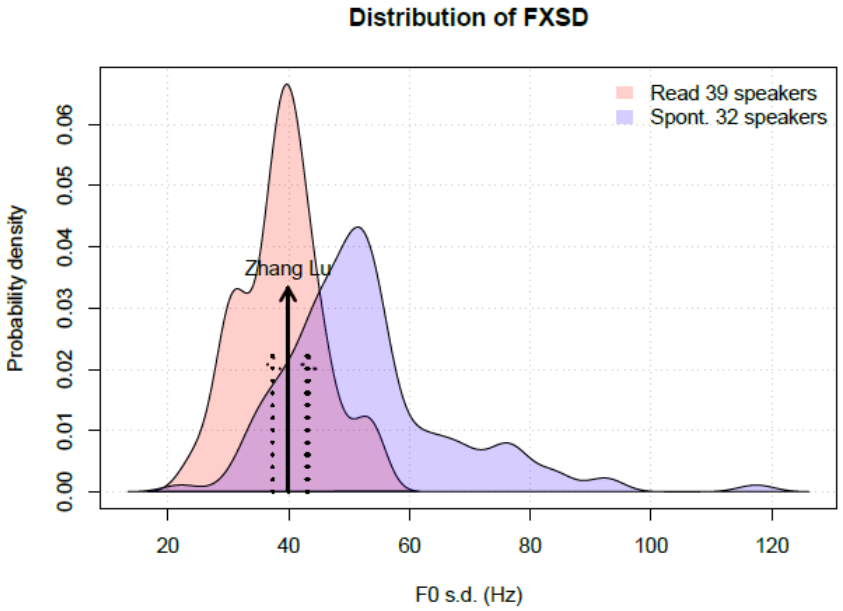



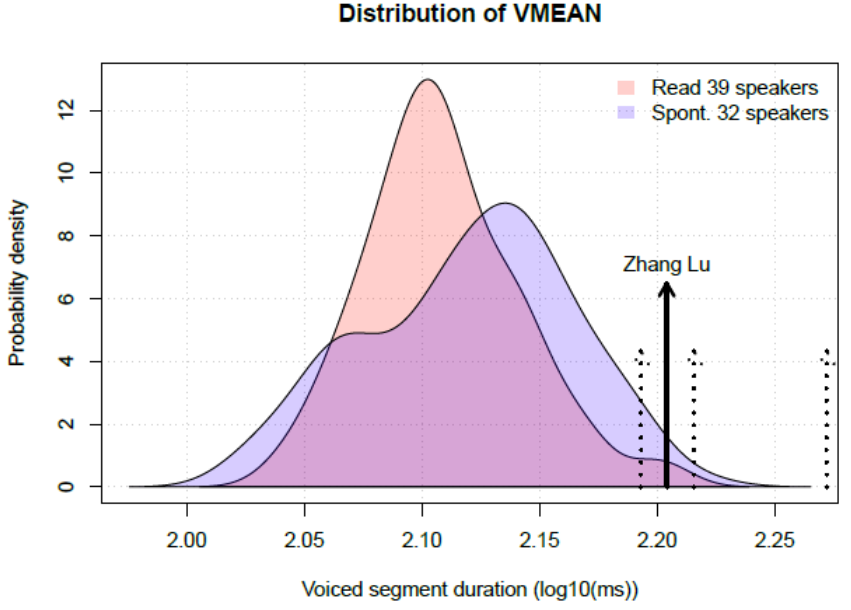



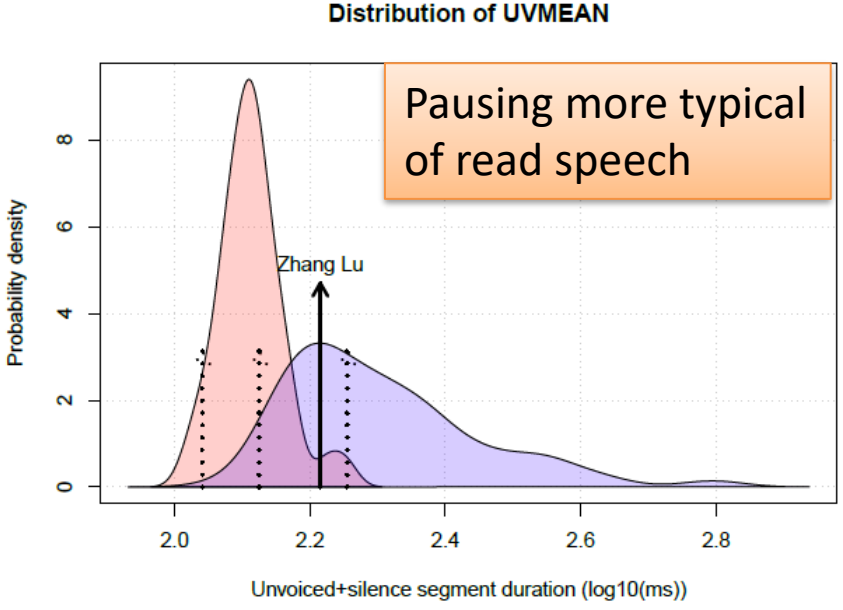

Speaking rate slower than reference speakers




Distribution of FXMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0.000
0.005
0.010
0.015
0.020
140
160
180
200
220
240
260
280
F0 mean (Hz)


Distribution of FXSD
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
20
40
60
80
100
120
F0 s.d. (Hz)


Distribution of VMEAN
Read 39 speakers
Spont. 32 speakers
Zhang Lu
Probability density
0
2
4
6
8
10
12
2.00
2.05
2.10
2.15
2.20
2.25
Voiced segment duration (log10(ms))


Distribution of UVMEAN
Pausing more typical of read speech
Zhang Lu
Probability density
0
2
4
6
8
2.0
2.2
2.4
2.6
2.8
Unvoiced+silence segment duration (log10(ms))

# Case Study Summary

- This pleasant voice:
  - Slightly raised pitch height - not high level of emotion
  - Low pitch variability – neutral emotion
  - Slow speaking rate – listener oriented speech
  - Few long pauses – fluent
- Also:
  - effective voice quality
  - good articulatory quality

# Conclusions

- Pleasant voice is easy to state but hard to define
- We present a framework based on independence of Linguistic, Extralinguistic and Paralinguistic properties
- Recent work in experimental paralinguistics shows we can relate listener ratings to acoustic properties
- We present a case study showing how acoustic properties of the voice can have paralinguistic interpretation

# Future Work

- We do not claim to have captured ‘pleasant voice’
- Merely that pleasant voice is amenable to quantification along with other paralinguistic properties
- A future study should collect a labelled corpus of student performances and listener ratings
- To learn associations that can be used to help learners