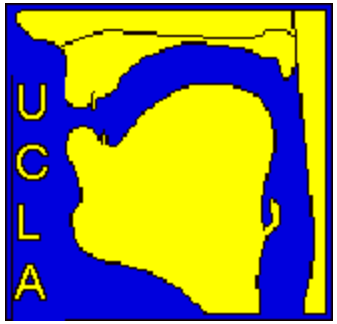
Voice quality variation with fundamental frequency in English and Mandarin

Patricia Keating

Phonetics Lab,

Linguistics, UCLA

keating@humnet.ucla.edu



Yen-Liang Shue

Speech Processing and

Auditory Perception Laboratory

Department of Electrical Engineering,

UCLA

Introduction

- Is voice quality related to voice pitch?
- Previous research has suggested that it is, both **across** and **within** speakers
- How are these patterns related? We look at relations **across** and **within** a single set of speakers

Across-speaker relation: Iseli, Shue & Alwan (2007)

- Speech samples from CID database (Miller et al. 1996)
  - 38 men, 37 women, 260 children: all American English speakers
  - Steady parts of several English vowels in real words in a carrier sentence
- Each token represented by one set of measurements
  - F0: ESPS method in Snack (Sjolander 2004)
  - H1*-H2*: amplitude difference between first 2 harmonics, corrected for formant frequencies and bandwidths
  - H1*-A3*: amplitude difference between corrected values of H1 and the third formant peak

Iseli et al. results: wedge-shaped relation

- F0 is positively related to H1*-H2* across low-pitched speakers

  (r = .767 for men)

- F0 is negatively related to H1*-H2* for speakers with F0 above about 175 Hz

  (r = -.47 for women and children)

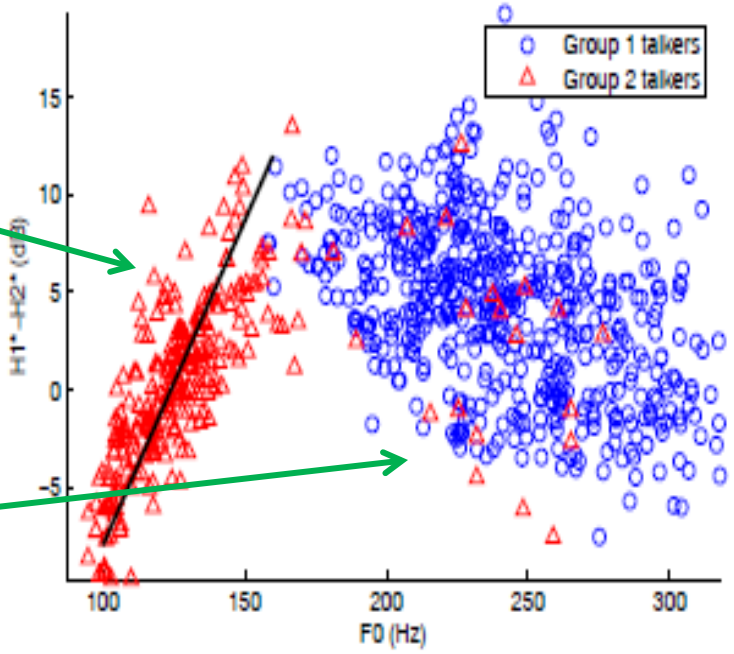


Fig. 5. Relation between $H_1^* - H_2^*$ and F_0 for high-pitched (group 1) talkers and low-pitched (group 2 talkers). A linear relationship for F_0 between 80 and 175 Hz is observed: see Eq. 1.

Summary: Previous ***across***-speaker relation

- Up to about 175 Hz, speakers who have overall higher-pitched voices generally have overall higher values of H1*-H2* than speakers who have overall lower-pitched voices
- Over 175 Hz, the opposite pattern holds, but less strongly
- Relation of F0 to H1*-A3* across speakers is much weaker, but also non-linear

Within-speaker relation: Swerts & Veldhuis (2001)

- Speech samples
  - 7 male Dutch speakers
  - /a/ with four different intonation (F0) contours
- Each speaker provides about 150 pairs of measurements from each utterance
  - F0 from inverse filtered signal
  - H1-H2 from inverse filtered signal
  - (also LF model parameters, not considered here)

Swerts & Veldhuis results

- F0 is often, but not always, positively related to source H1-H2 within individual speakers
- Table shows r values:
  - ★ marks positive highly-significant correlations

Table 2
Correlations between F_0 and H1–H2 in different intonation patterns for different speakers

SP	Intonation pattern			
	hl	hlh	lh	lhl
DH	0.12	★ 0.60[a]	★ 0.47[a]	0.02
EK	★ 0.65[a]	★ 0.50[a]	★ 0.67[a]	★ 0.70[a]
GV	−0.36[a]	−0.56[a]	−0.45[a]	−0.16[a]
JP	★ 0.82[a]	0.14[b]	−0.57[a]	★ 0.65[a]
MS	★ 0.81[a]	★ 0.78[a]	★ 0.92[a]	★ 0.70[a]
RK	0.16	★ 0.73[a]	★ 0.42[a]	0.26[b]
RH	−0.22[b]	0.18[c]	0.24[b]	0.13

[a] $p < 0.001$ ★
[b] $p < 0.01$.
[c] $p < 0.05$.

Summary: Previous ***within***-speaker relation

As an individual man's F0 goes up, his (source) H1-H2 generally also goes up (i.e. more prominent H1)

Present study

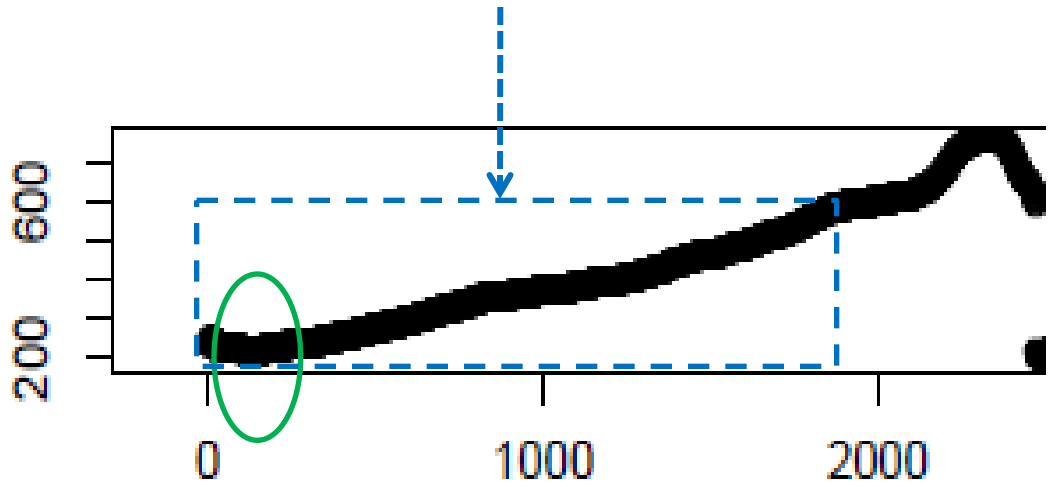
- Same speakers and utterances for within- and across-speaker comparisons
- Males and females
- Two languages
- Additional acoustic voice measures

Speech samples

- Repeated rising and falling tone sweeps on [a]
- Speakers began at self-selected comfortable pitch
- Swept up or down in pitch to their highest or lowest comfortable pitch
- Swept down into creaky voice
- Each sweep about 2-5 sec long

Speech samples

- **Beginnings** of sweeps (2nd –10th percent) tested for **across**-speaker relations
- **Almost entire** sweeps tested for **within**-speaker relations – up to 600 Hz for females, up to 500 Hz for males (so that F0 is below F1)



Speakers

- Mostly UCLA students
- Native Mandarin speakers were mostly from Taiwan, and all spoke English
- 46 recorded in total; 5 could not be used here

INCLUDED	men	women
English	10	10
Mandarin	11	10

Acoustic analysis

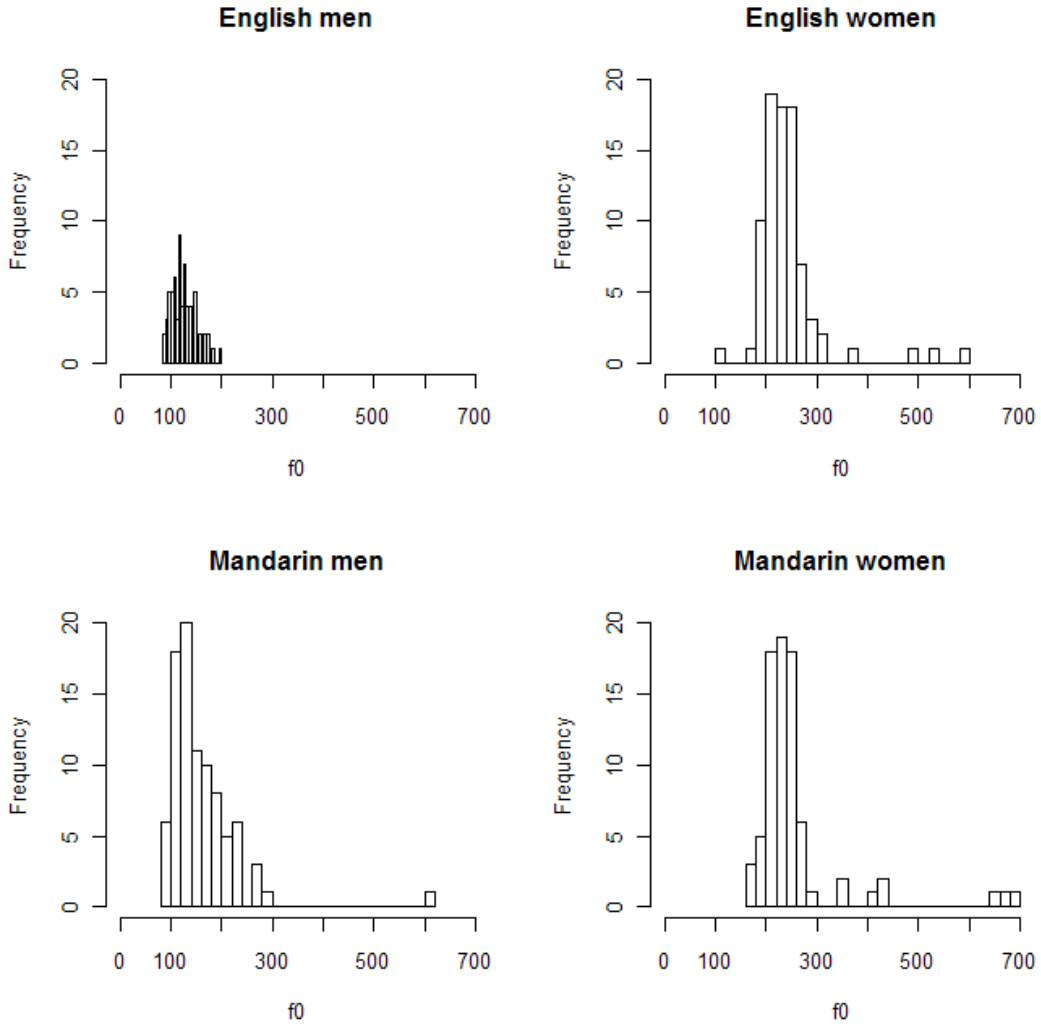
From VoiceSauce, a new program for voice analysis (Shue et al. 2009):

- F0 by the STRAIGHT algorithm (Kawahara et al. 1998)
- Energy, Cepstral Peak Prominence
- (formant frequencies and bandwidths)
- Corrected (shown with *) and uncorrected harmonic amplitude difference measures made from the audio signal (Iseli et al. 2007)

Harmonic amplitude measures

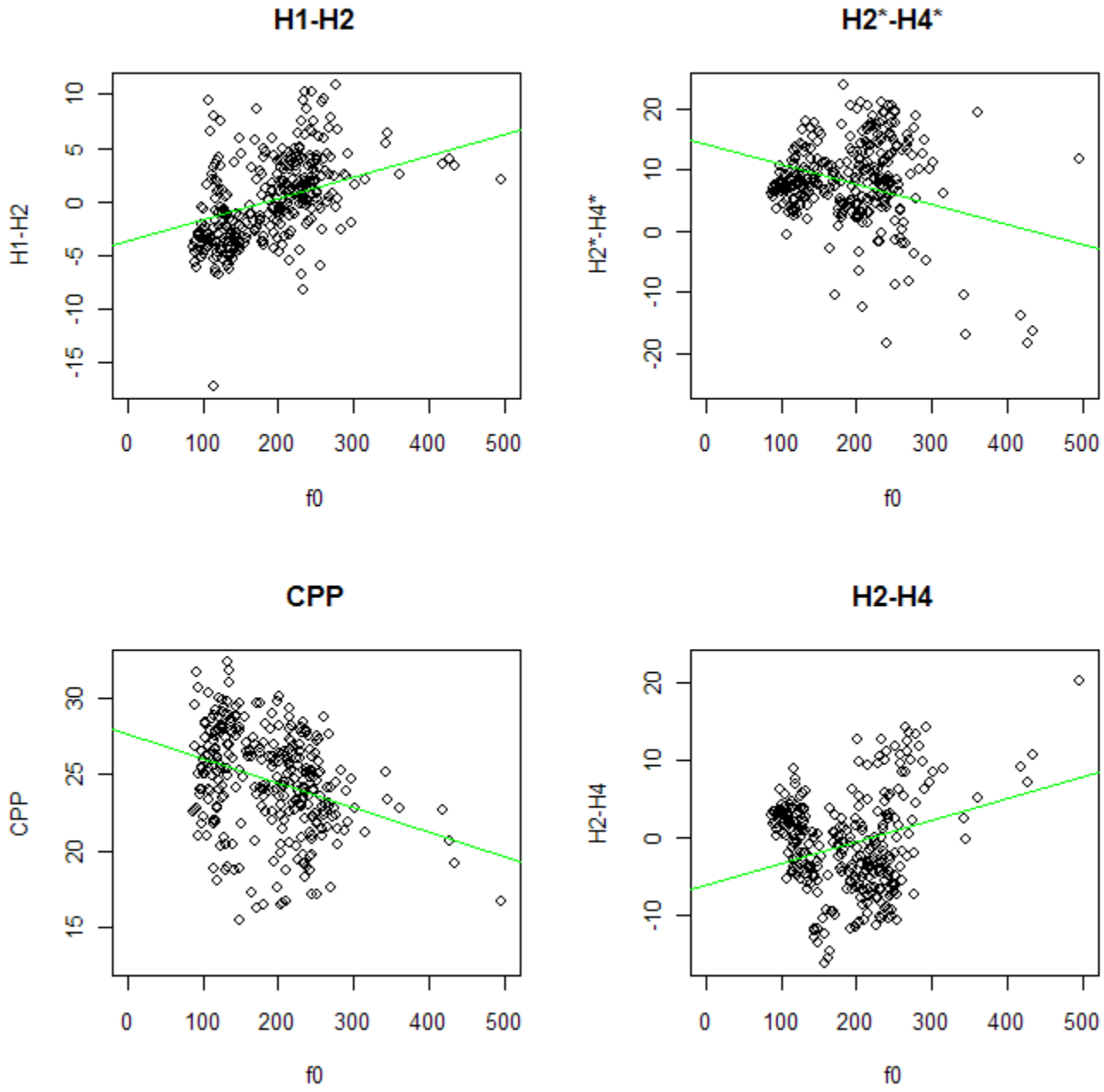
- H1-H2
- H2-H4
- H1-A1
- H1-A2
- H1-A3
- Same, but corrected for formant frequencies and bandwidths

Distributions of initial F0 values in across-speaker dataset



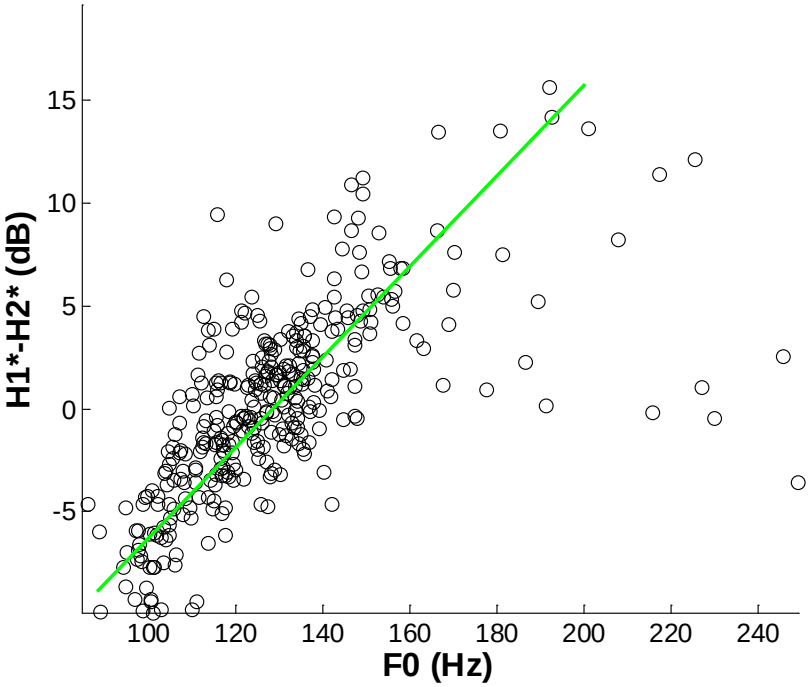
Results: Across all speakers

- Relatively few significant correlations that account for more than 10% of variance by linear regression
- Best overall correlations (see next slide):
  - Uncorrected H1-H2 (r = .45)
  - Cepstral Peak Prominence (r = -.41)
  - H2*-H4* (r = -.39)
  - (Uncorrected H2-H4 (r = .37)) (artifact of F0)
  - But NOT H1*-H2*, H1*-A3*

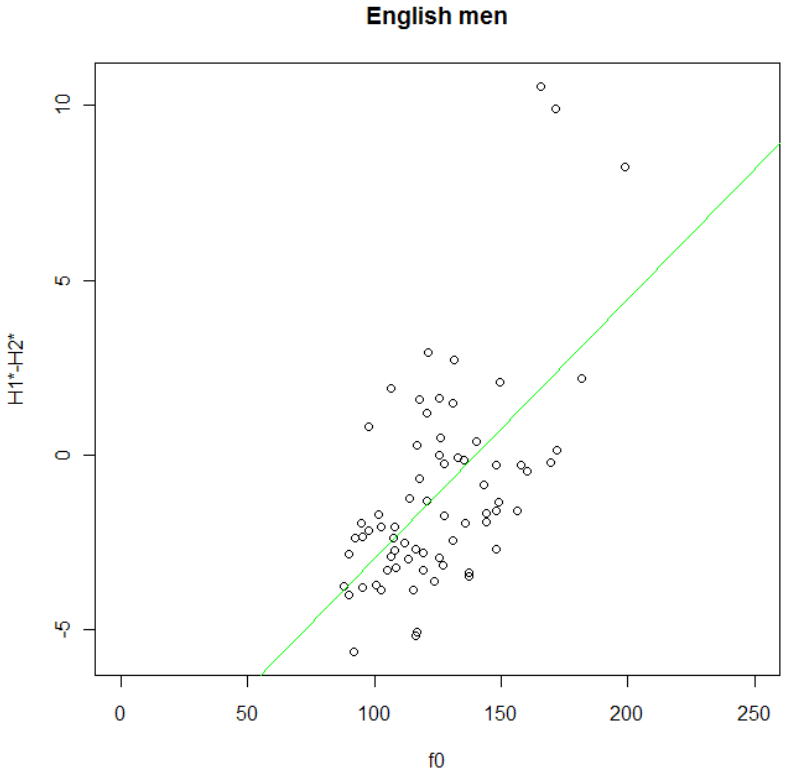

H1-H2
H1-H2
f0
H2*-H4*
H2*-H4*
f0
CPP
CPP
f0
H2-H4
H2-H4
f0

H1*-H2*: Comparison with Iseli et al. (English men only)

Iseli et al. (r = .77)

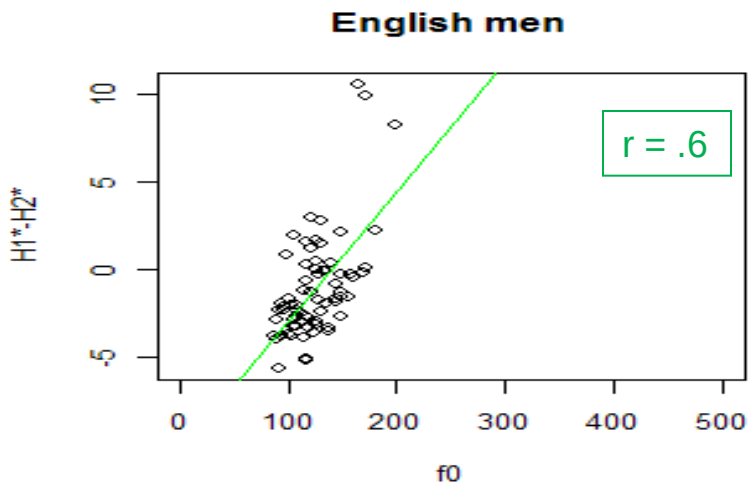


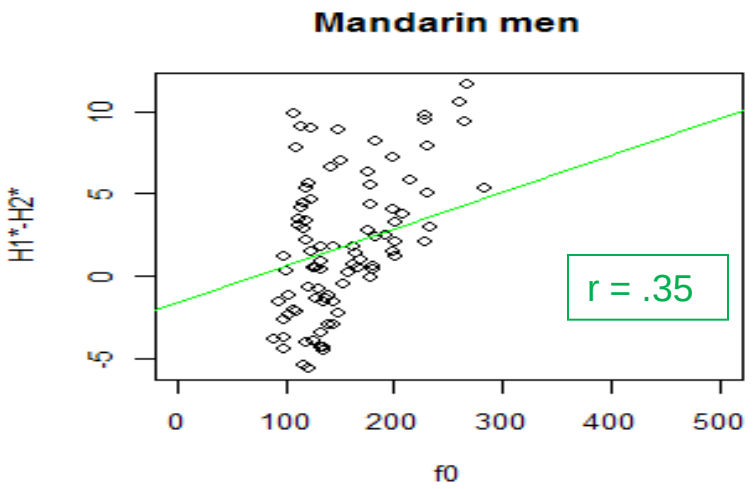
Present study (r = .6)

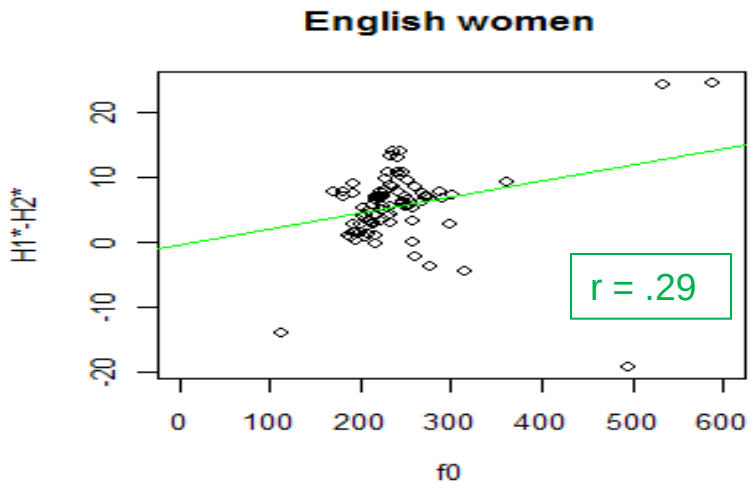


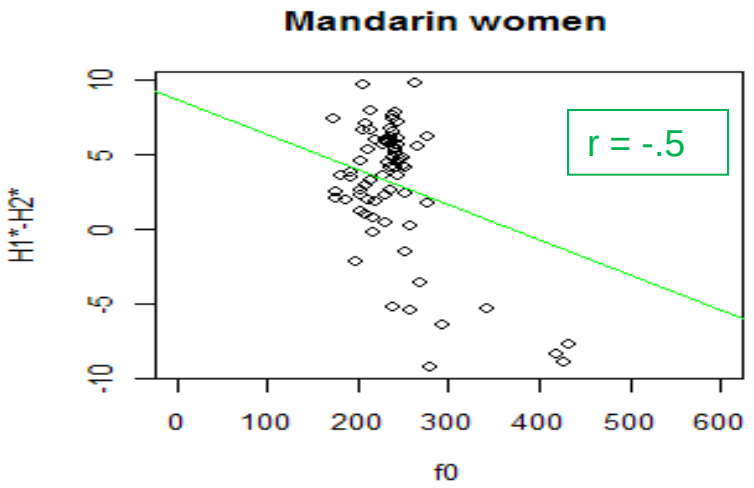
With higher F0, H1 is more prominent

Only Mandarin women show a negative relation for H1*-H2*

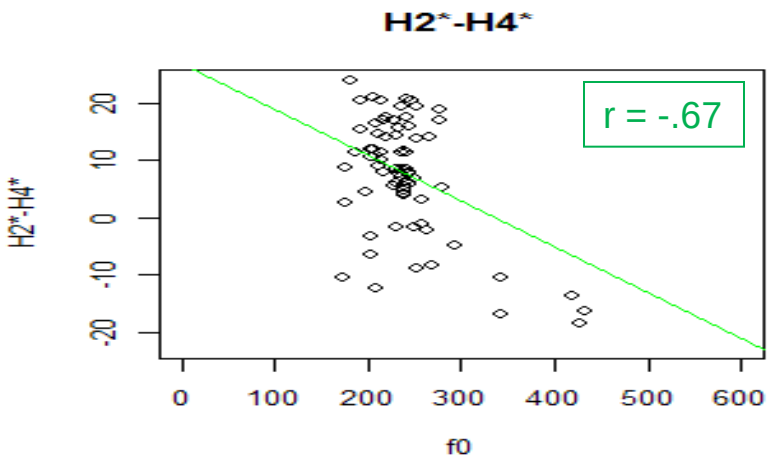


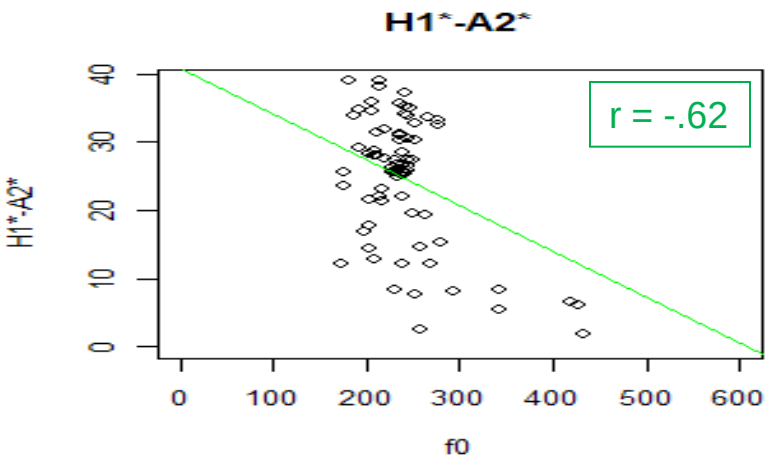


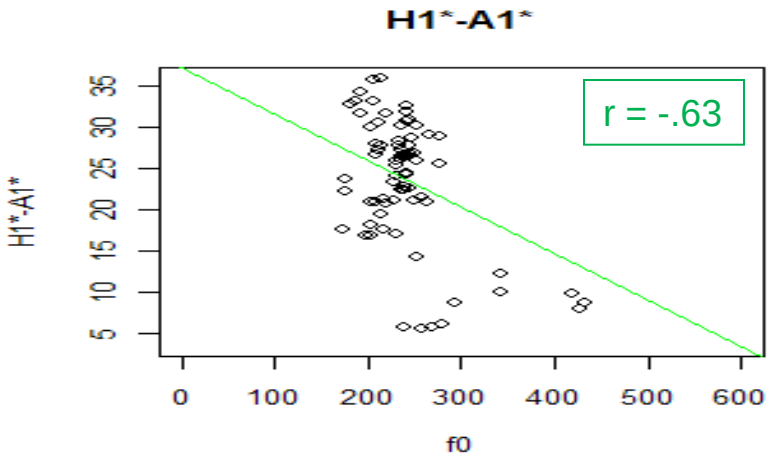


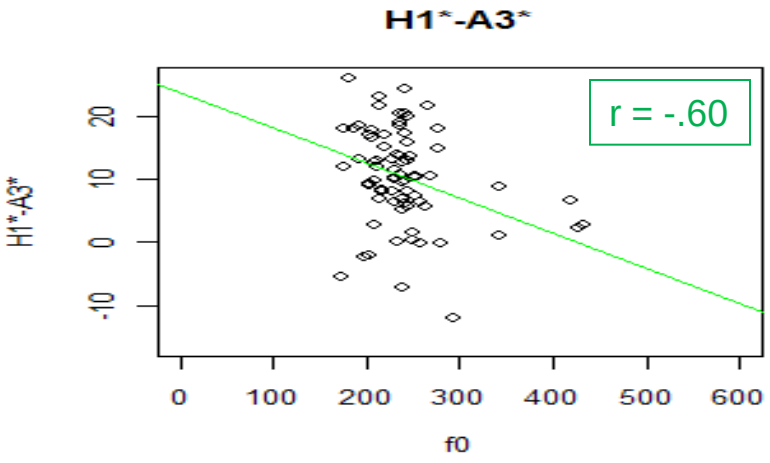


Mandarin women show other correlations with |r| >.5









Summary: Across-speaker correlations

- Relatively few correlations of even moderate strength, and none very strong
- Less of a non-linear relation across F0s than Iseli et al. (2007) found, with fewer high F0 values for our English men and no turning point for Mandarin men
- Relation of H2*-H4* to F0 is new finding, this correlation is higher in Mandarin
- Other differences between speaker groups that can't be attributed to F0 differences

Results: Within speakers

- Most correlations (of voice measures with F0) for individual utterances are significant
- All acoustic measures show significant correlations for most speakers

Most often significant is **H1*-H2***

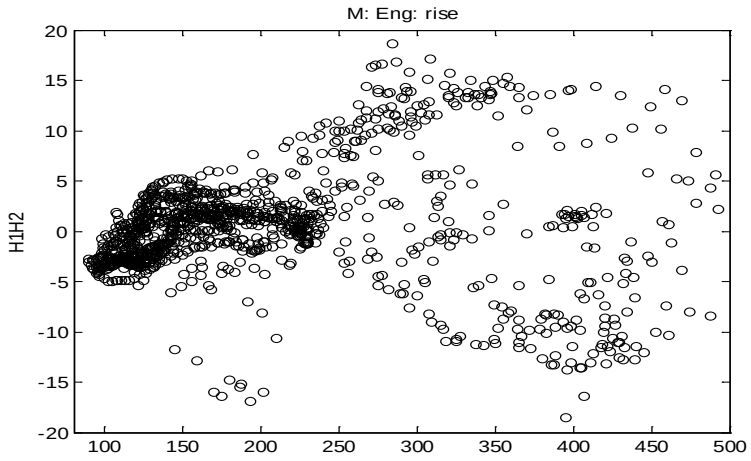
- In many utterances, F0 accounts for >50% of variance in H1*-H2*
- But we often see non-linear relations, though with a higher-F0 turning point than in the Iseli et al. wedge-shaped pattern
- Plots show datapoints over 10-90% of each utterance, for 2% intervals, *pooled across speakers and utterances*

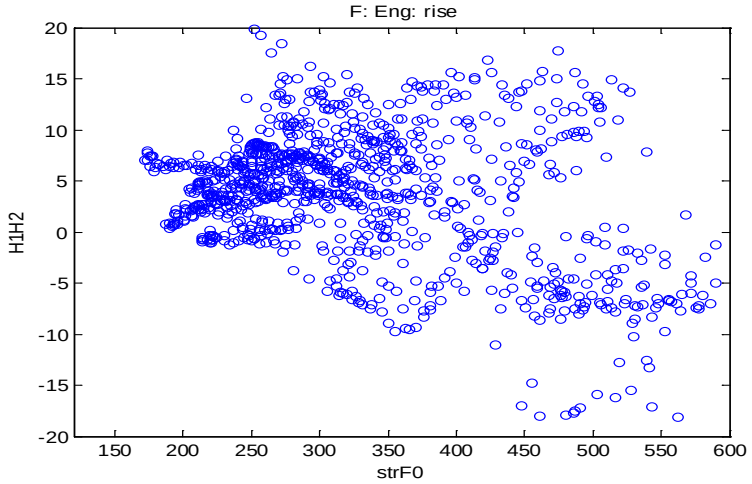
H1*-H2* pooled by subgroup for rising-F0 sweeps

Men

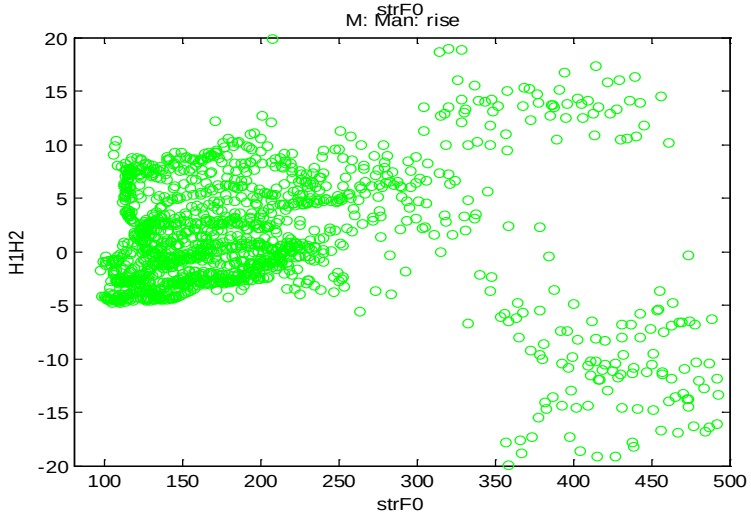
Women

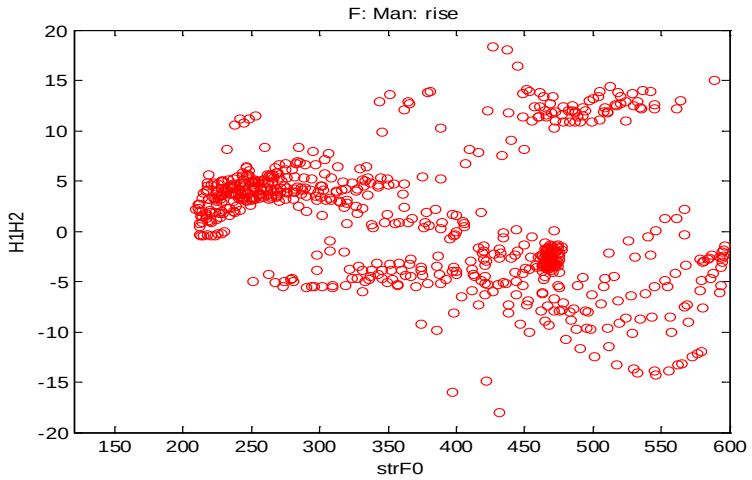
Eng



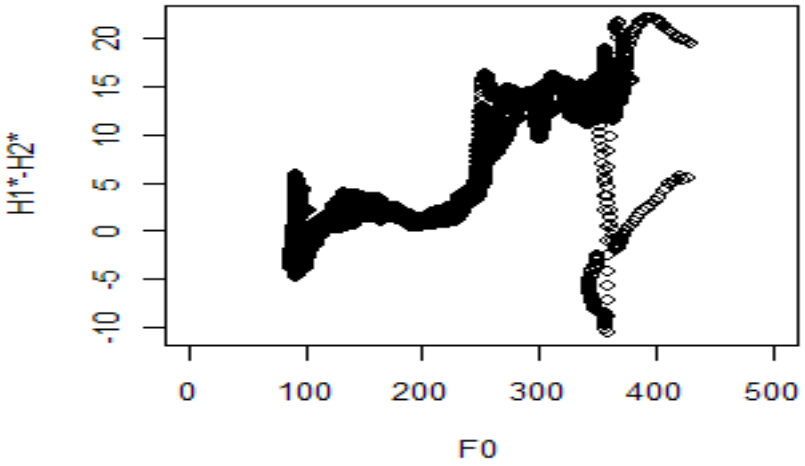


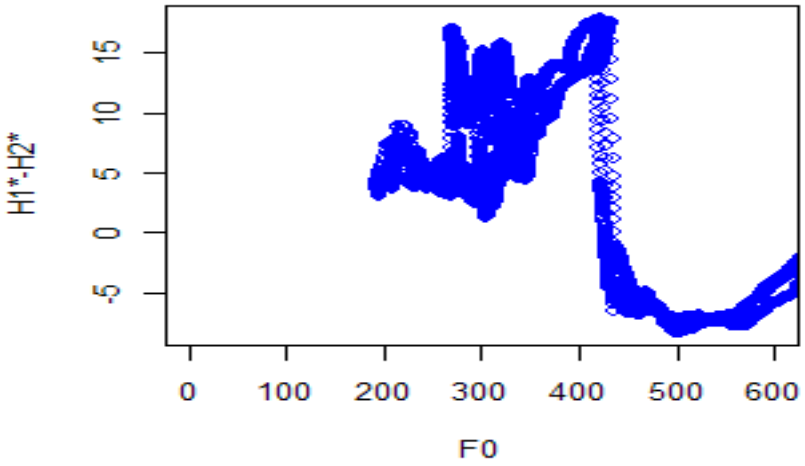
Mand

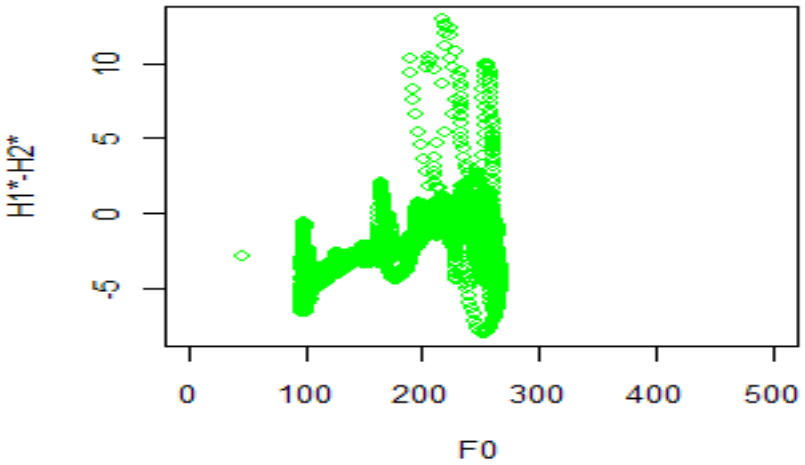


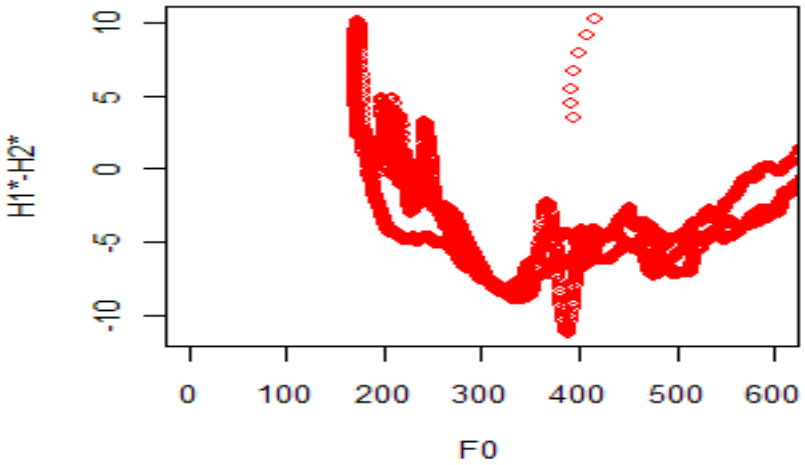


Some individual speakers







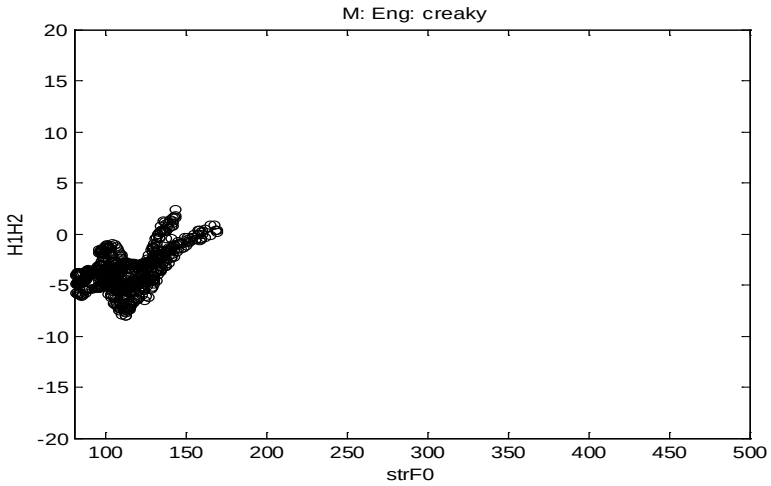


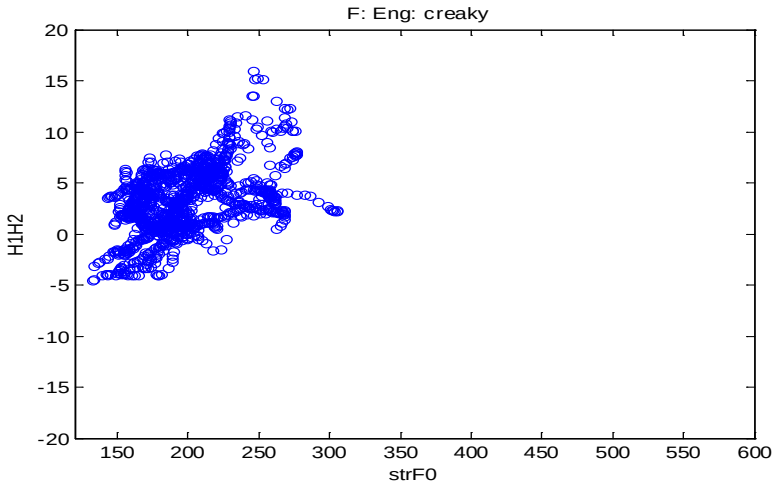
H1*-H2*: Clearly positive relation in low-F0 falls

Men

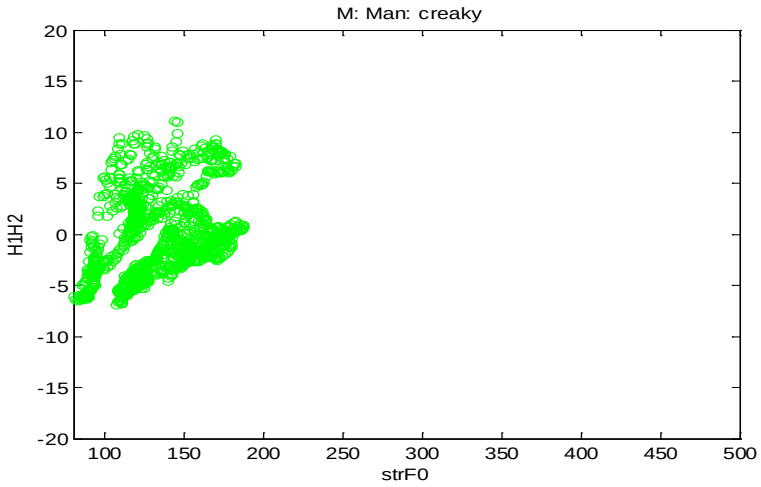
Women

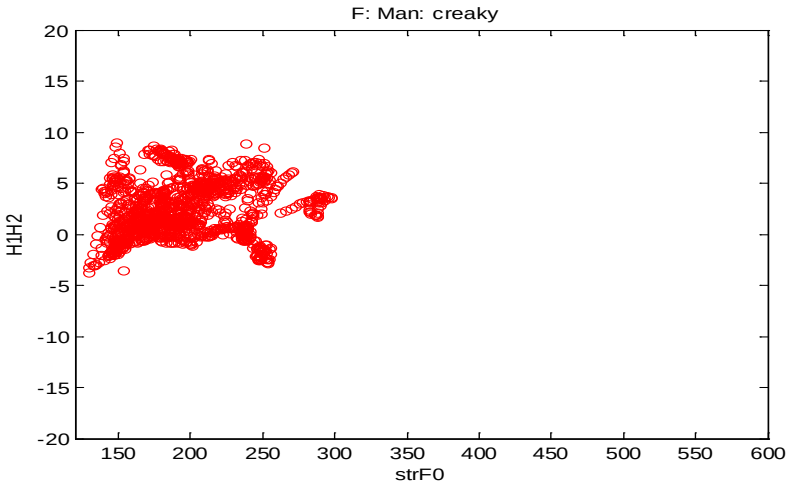
Eng





Mand





H1*-H2*: Comparison with Swerts & Veldhuis

- As in S&V, many strong correlations
- As in S&V, correlations more often positive, but can also be negative for some speakers
- Our range of F0 values is larger
  - In our data, correlations tend to be positive when F0 is low (in falls), but negative with higher F0s (in rises)

Other correlations within speakers

- Some within-speaker relations with F0 are consistently found across speakers:
  - H1*
  - H2*
  - Energy
- Other relations are almost always significant within speakers, but speakers differ in direction, with no patterns by speaker-subgroup or F0 range/contour

Comparison: Across- vs. within-speakers

- Few significant/strong correlations ***across*** all speakers, but many significant/strong correlations ***within*** each speaker, though hugely variable
- The wedge-shaped relation of H1*-H2* to F0 found by Iseli et al. ***across*** speakers is more apparent in our ***within***-speaker data, with its wider F0 ranges; across speakers it is clearer in Mandarin than in English

Summary and conclusions

- Does voice quality differ systematically across speakers who have different comfortable voice F0s? Not much; most strongly, but non-linearly, in H1*-H2*
- Does voice quality change systematically as a speaker changes his/her pitch? Yes, along many dimensions, but often idiosyncratically; most consistently, but non-linearly, in H1*-H2*

References

- Iseli, M., Y.-L. Shue and A. Alwan (2007) “Age, sex, and vowel dependencies of acoustic measures related to the voice source”, *J. Acoust. Soc. Am.121*, 2283-2295
- Kawahara, H., A. de Cheveign and R. D. Patterson (1998) “An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMPO in the STRAIGHT-suite,” in *Proceedings ICSLP’98*, Sydney, Australia, December 1998
- Miller, J., S. Lee, R. Uchanski, A. Heidbreder, B. Richman and J. Tadlock (1996) “Creation of two children’s speech databases”, in *Proceedings of ICASSP*, Vol. 2, pp. 849-852
- Shue, Y.-L., P. Keating and C. Vicenik (2009) “VoiceSauce: A program for voice analysis”, poster 2pSC2 at this meeting
- Sjolander, K. (2004) "Snack sound toolkit," KTH Stockholm, Sweden. http://www.speech.kth.se/snack
- Swerts, M. and R. Veldhuis (2001) “The effect of speech melody on voice quality”, *Sp. Comm. 33*, 297-303

Acknowledgments

- NSF grant BCS-0720304; UCLA Committee on Research grant
- Co-PIs: Abeer Alwan and Jody Kreiman
- Grace Kuo for help with data analysis
- Caitlin Smith and Ting Fang for making audio recordings