

Reexamining cue enhancement: The case of whispered tones in Mandarin

Charles B. Chang <cbchang@berkeley.edu>

Yao Yao <yaoyao@berkeley.edu>

University of California, Berkeley

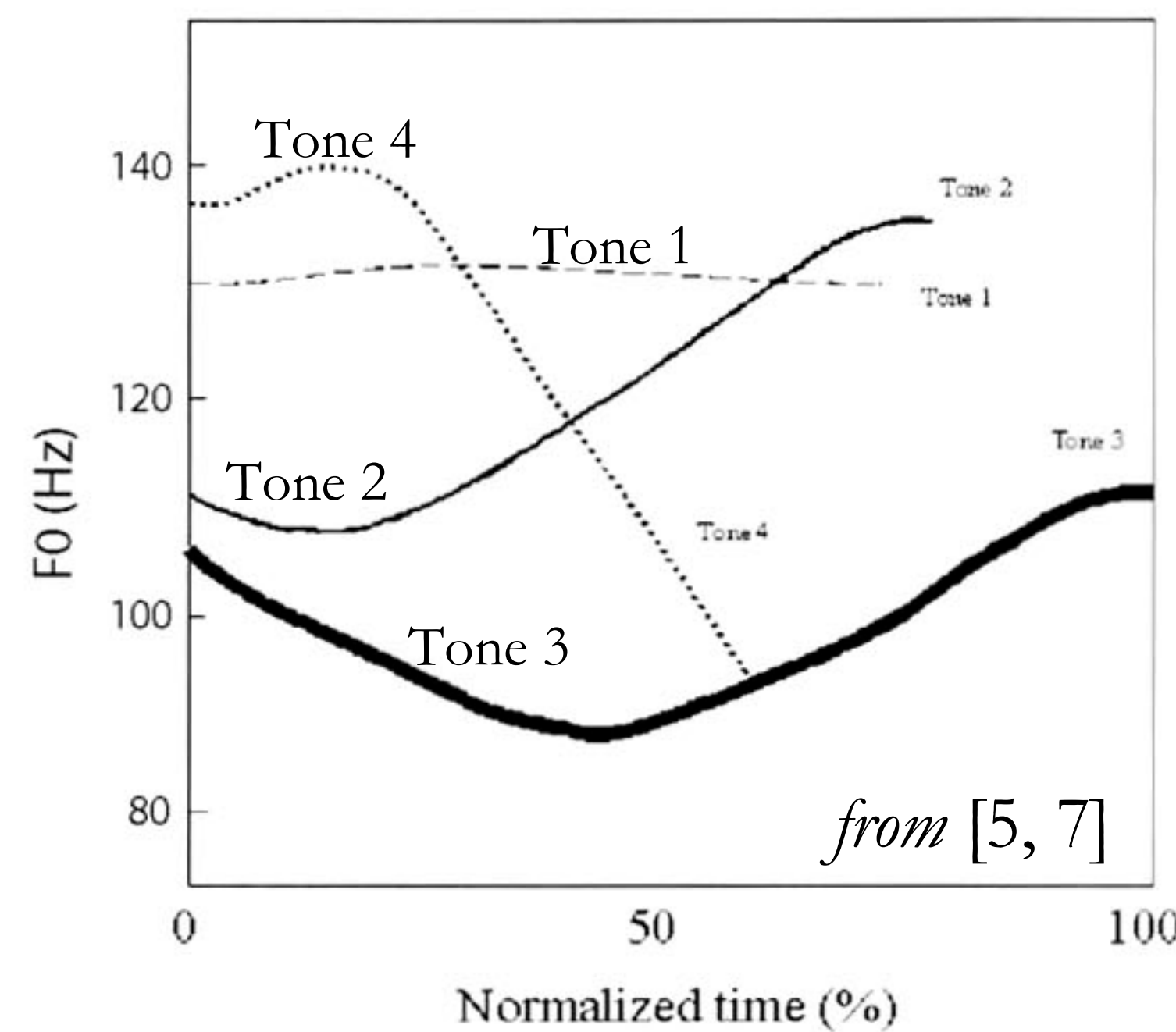
Department of Linguistics



1. Introduction

➤ 4 basic lexical tones in Mandarin

- Tone 1: high level (55)
- Tone 2: mid rise (25)
- Tone 3: low fall-rise (214)
- Tone 4: high fall (51)



➤ Differences in **F0**, **duration**, and **intensity**

2. Background

- How well are tones recognized in whisper (which has neither F0 nor harmonic fine structure)?
 - 60-80% accuracy in Mandarin [4]
 - poorer accuracy in Vietnamese, which has more tones [6]
- What secondary cues can speakers attend to in whisper?
 - **temporal envelope / intensity contour** [2, 3, 8, etc.]
 - **syllable duration** [5]
- [5]: better recognition accuracy with “human whispered” stimuli than “machine whispered” stimuli (signal-processed to remove F0)
 - [5]’s conclusion: in whisper, talkers promote secondary cues since they know the primary cue (F0) is not available to the listener
- **Research question #1: How do the four Mandarin tones compare acoustically in whisper as compared to normal speech?**
- **Research question #2: Do talkers exaggerate secondary cues to tone in whisper to make up for the absence of F0 information?**

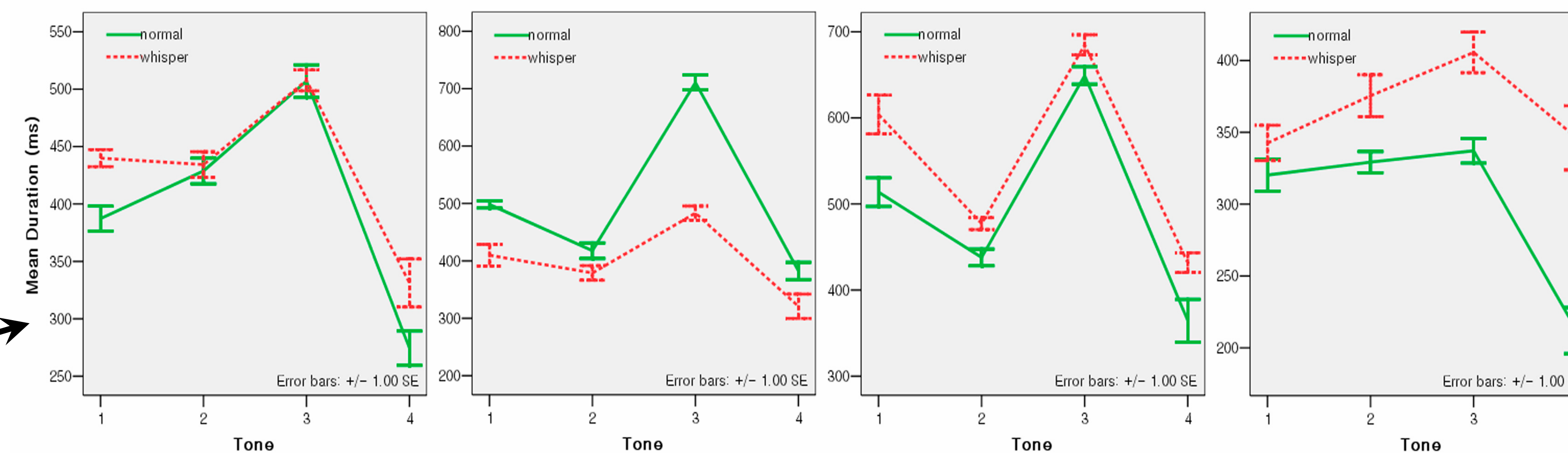
2. Methods

- **Materials:** list of 30 minimal tone quadruplets (= total of **120 test items**)
 - open syllables with unaspirated stop or fricative onsets
- **Subjects:** 4 native Mandarin speakers from mainland China (2 m, 2 f)
 - 3 from Beijing, 1 from Shanghai
 - all in their 20s-30s, no history of articulatory or auditory problems
- **Recording:** speech recorded digitally in quiet rooms at 44.1 kHz / 16 bps
 - Marantz PMD670 solid state recorder, AKG C420 head-mounted condenser mic positioned to the side of the mouth about 2 cm away
 - stimuli recorded in isolation in random order for each speech genre
 - 120 items/genre x 3 tokens/item = 360 tokens for 1 speaker/genre
- **Measurements:** taken in Praat 4.5.14 [1] on a Fourier spectrogram (Gaussian window, length 5 ms, dyn. range 70 dB, pre-emp. 6 dB/oct)
 - **vowel duration** measured from the end of the onset burst / strident interval to the end of visible formant structure in the vowel
 - **average intensity** measured over the same vowel interval

3. Results

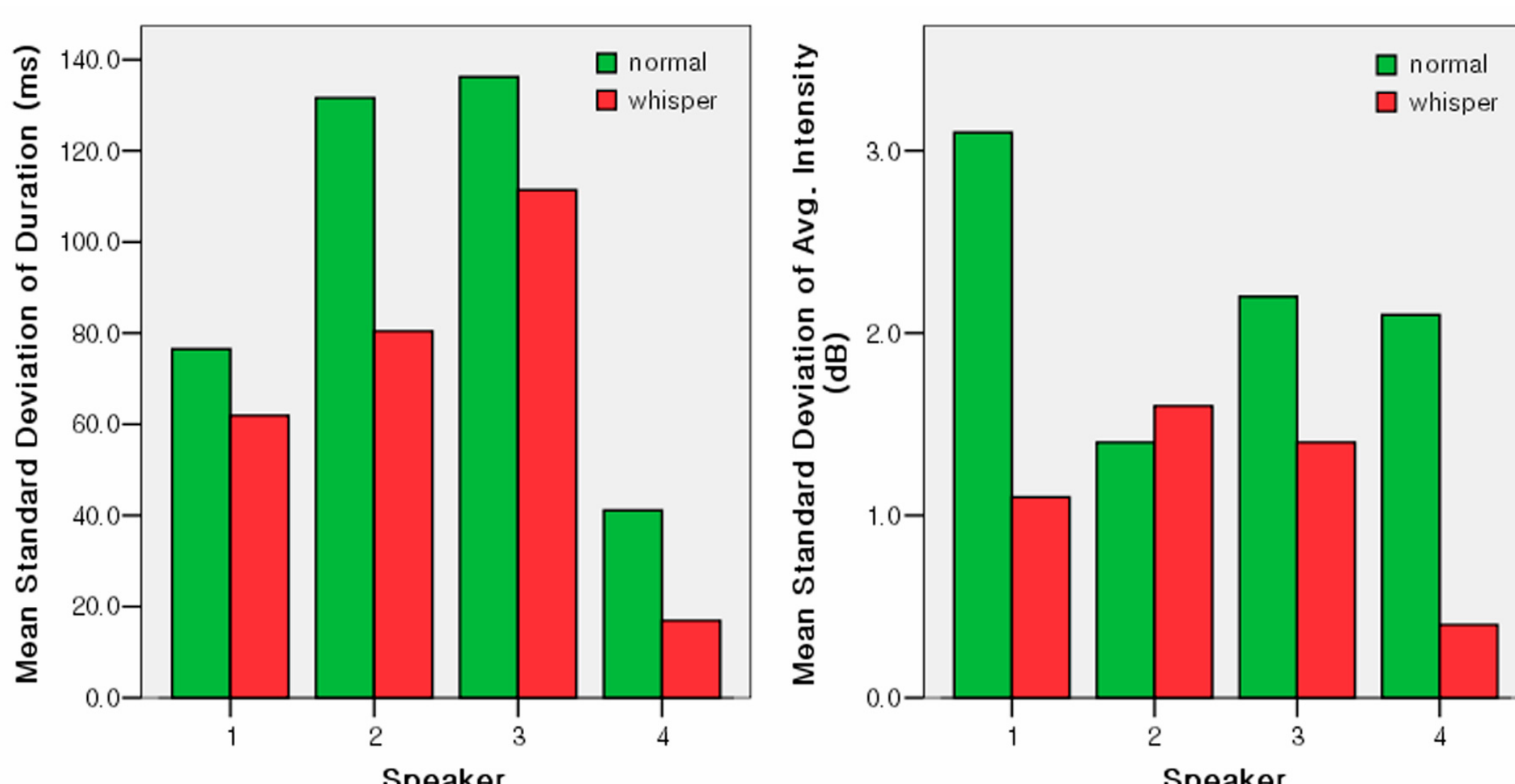
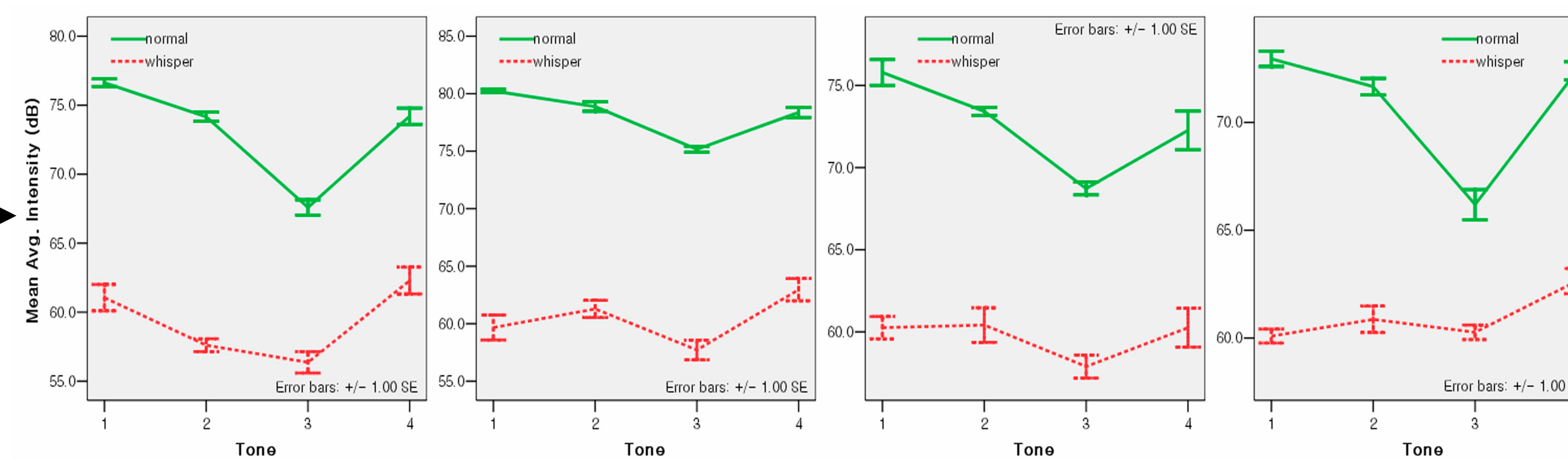
➤ **Duration:** same hierarchy among tones in normal speech and whisper: Tone 3 > Tone 1/Tone 2 > Tone 4

- duration differences in normal speech significant at $p < .05$ except Tone 1 vs. Tone 2 for Sp. 1, 2
- duration differences in whisper significant at $p < .05$ except Tone 2 vs. Tone 3 for Sp. 2
- **main effect of tone** for all speakers (L to R: Sp. 1-4)
- **main effect of speech genre** for Sp. 1, 3, 4
- [tone x speech genre] interaction for all speakers



➤ **Intensity:** similar hierarchies in both speech genres

- normal speech: Tone 1 > Tone 2 > Tone 4 > Tone 3
- whisper: Tone 4 > Tone 1, Tone 2, Tone 3
- **main effect of tone** for all speakers (L to R: Sp. 1-4)
- **main effect of speech genre** for all speakers
- [tone x speech genre] interaction for all speakers



➤ Standard deviations (SDs) show **less variability in whisper**

- **duration:** smaller SDs for all speakers
- **intensity:** smaller SDs for Sp. 1, 3, 4; little diff. for Sp. 2

- Presence of **inter-speaker variability** in duration/intensity patterns in whisper
 - **duration:** Sp. 1, 3, and 4 lengthen, but Sp. 2 shortens tones in whisper
 - **intensity:** Sp. 1 orders tones as 4 > 1 > 2 > 3; Sp. 2/3 order as 4 > 2 > 1 > 3; Sp. 4 orders as 4 > 2 > 3 > 1

4. Conclusions

- There are **significant differences in duration and intensity** among the 4 Mandarin tones in both normal and whispered speech
 - relative differences are similar across speech genres
- Talkers are **not exaggerating secondary cues** in whispered speech
 - duration/intensity differences among tones become smaller in whispered speech (contra [5])
 - [5]’s perception results are probably due to unnaturalness of the “machine whispered” stimuli rather than exaggerated cues in the “human whispered” stimuli
- There is significant **individual variation in the production of whispered tones**, which may be correlated with gender and/or dialect

Acknowledgments U.S. Department of Education, UC Berkeley Phonetics & Phonology Forum, International Congress of Phonetic Sciences, all speaker participants

References

- [1] Boersma, P., Weenink, D. 2007. Praat: doing phonetics by computer. <http://www.praat.org>.
- [2] Fu, Q.-J., Zeng, F.-G. 2000. Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language, and Hearing* 5: 45-57.
- [3] Fu, Q.-J., Zeng, F.-G., Shannon, R.V., Soli, S.D. 1998. Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America* 104: 505-510.
- [4] Liang, Z.-A. 1963. Hanyu putonghua zhong shengdiao de tingjiao bianren yiju (The auditory basis of tone recognition in Standard Chinese). *Acta Phys. Sin.* 26: 85-91.
- [5] Liu, S., Samuel, A. 2004. Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech* 47(2): 109-138.
- [6] Miller, J.D. 1961. Word tone recognition in Vietnamese whispered speech. *Word* 17: 11-15.
- [7] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.
- [8] Xu, L., Tsai, Y., Pfingst, B.E. 2002. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *Journal of the Acoustical Society of America* 112: 247-258.