Relative role of pitch vs. phonation cues in White Hmong tone identification

Marc Garellek[a], Christina M. Esposito[b], Patricia Keating[a], & Jody Kreiman[c]

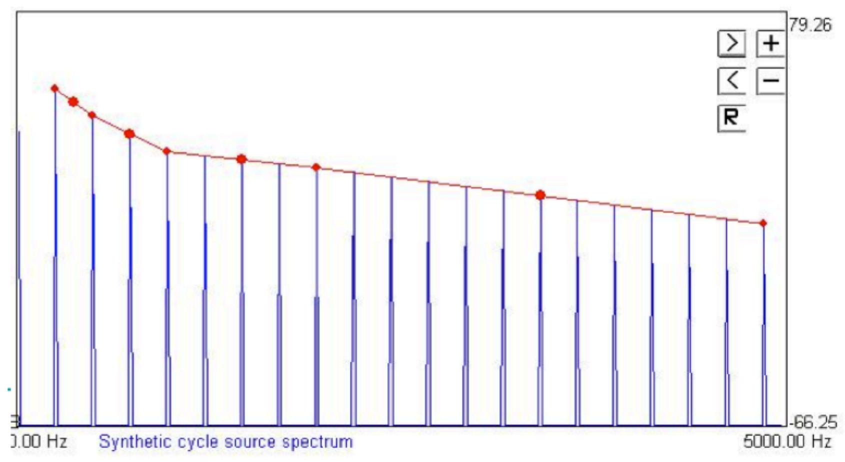
[a]Department of Linguistics, UCLA; [b]Department of Linguistics, Macalester College;
[c]Department of Head and Neck Surgery, UCLA

marcgarellek@ucla.edu

Introduction

- Source spectrum often schematized with harmonic amplitudes falling off at 12 dB/octave (Ní Chasaide & Gobl 1997).
  - But, there is much variation across voices and voice qualities (see Fig.1)
- ***How can we model the source spectrum to capture differences across speakers & voice quality types?***
- Kreiman et al. (2011) model the source spectrum using 4 component slopes:
  - H1-H2, H2-H4, H4-2kHz, 2kHz-5kHz

Fig 1



- Model was found to be good fit to cross-speaker variation.
  - Unclear whether these components are perceptually relevant.
- ***What components of source spectrum are used to perceive non-modal phonation?***
- We use a language with non-modal phonation as a test case.
- In White Hmong, breathy and creaky voice accompany certain lexical tones:
  - High-falling breathy (52) tone (cf. high-falling modal (52) tone).
  - Low-falling creaky (21) tone (cf. low modal (22) tone, which is longer in duration: Esposito 2012).
- First, we test whether breathy and creaky voice are contrastive with modal.

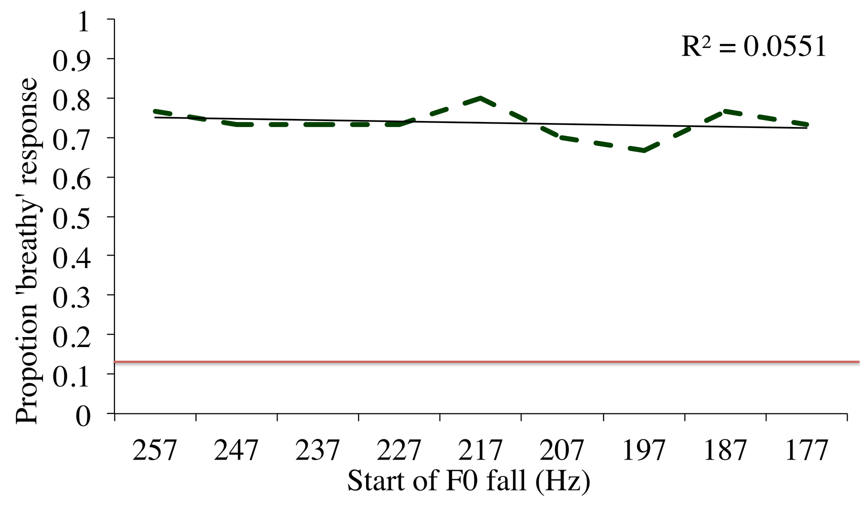
Experiment 1

- ***Do breathiness and creakiness contrast with modal voice in Hmong?***
- Stimuli were monosyllables from original 7 tones, but with f0 resynthesized using PSOLA (Moulines & Charpentier 1990).
  - Voice quality is preserved.
  - E.g. instead of high-falling breathy (52) tone, f0 was low-falling (21).
  - Also manipulated duration for comparing low-falling creaky (21) with low modal (22).
- 15 Hmong listeners participated in a word identification task:
  - Decide which word they heard from among 7 orthographic choices.
  - 7 alternatives = 7 productive tones.

RESULTS

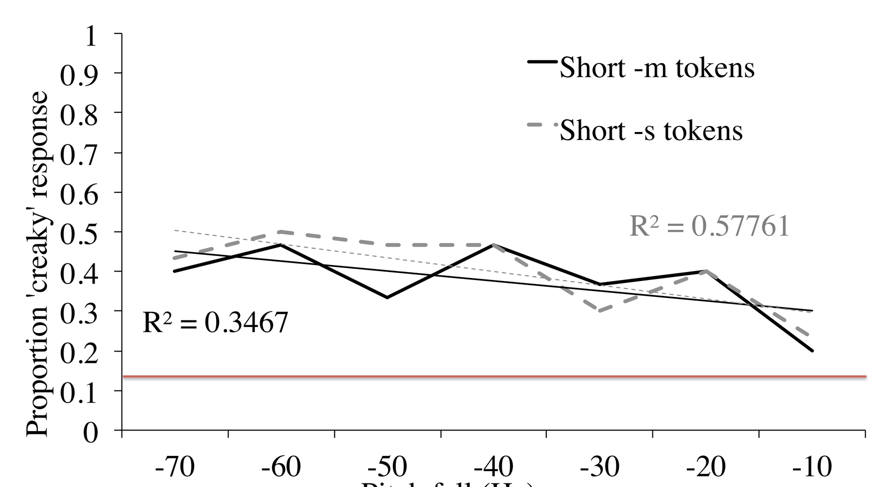
- For a token to be heard as breathy, the original stimulus had to be breathy (f0 <u>not</u> significant!)

Fig 2



- For a token to be heard as creaky, original stimulus could be modal, as long as f0 was low-falling and duration was short (phonation <u>not</u> significant!)

Fig 3



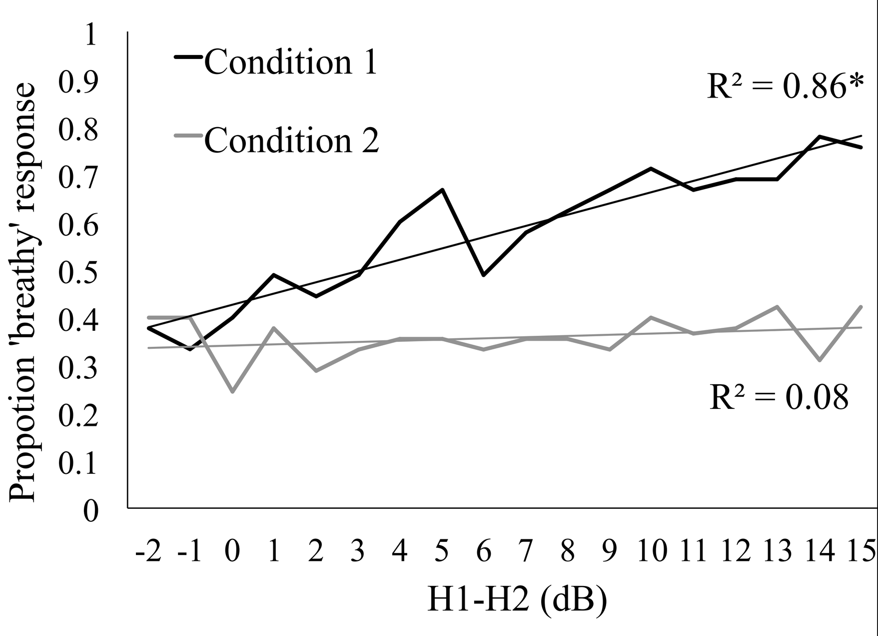
Experiment 2

- ***Breathiness is contrastive in Hmong, but what in the source spectrum cues breathiness?***
- Same listeners as in Expt 1:
  - Decide which word they heard.
  - Had to choose between breathy vs. modal word.
- Stimuli used same high-falling f0 contour, manipulate slopes of the source spectrum model (Kreiman et al. 2011):
  - Condition 1: H1-H2 varied from -2 to 15 dB, all other components held constant.
  - Condition 2: As H1-H2 increased, H2-H4 decreased linearly (from 22 to 8 dB).
  - Other conditions for assessing importance of H4-2kHz and 2kHz-5kHz were included.

RESULTS

- All components, but esp. H1-H2 and H2-H4, are significant in predicting breathy responses.
- ***H1-H2 and H2-H4 are independently important in predicting breathy responses:***
  - If H1-H2 is very low but H2-H4 high → modal percept.
  - If H1-H2 is very high but H2-H4 is low → modal percept.

Fig 4



Discussion

- For a given language, non-modal phonation can be used contrastively or to enhance a pitch contrast:
  - Breathy voice in Hmong is the main cue to contrast between two high falling tones.
  - Creaky voice enhances low f0 and short duration, but not necessary to identifying the low-falling (21) tone.
- The current model of the source spectrum can adequately capture contrastive breathy vs. modal voice in Hmong
  - Listeners prefer sharp decreases in harmonic amplitude between H1 and H4 to perceive breathiness.
- Higher-frequency energy is much less important in cueing breathy voice.
- Although inharmonic energy is often cited as an important cue to breathiness, here breathy vs. modal voice was distinguished only by harmonic energy.
- What articulatorily is responsible for changes in H2-H4?
  - Still unclear, but current work suggests vocal fold stiffness and asymmetry (Zhang et al. 2011).

Acknowledgments

Thanks to the Hmong-American Partnership in St. Paul, MN, and to Norma Antoñanzas-Barroso. This work is supported by NSF grants BCS-0720304 and IIS-1018863, and NIH/NIDCD grant DC01797.

References

1. Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong phonation. *JPhon* 40, 466-476.
2. Kreiman, J., Garellek, M., & Esposito, C. (2011). Perceptual importance of the voice source spectrum from H2 to 2 kHz. *JASA* 130, 2570.
3. Moulines, E. & Charpentier, F. (1990). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *SpCom* 9, 453–467.
4. Ní Chasaide, A. & Gobl, C. (1997). Voice source variation. In *The Handbook of the Phonetic Sciences*, Oxford: Blackwell.
5. Zhang, Z., Kreiman, J. & Gerratt, B. R. (2011). Perceptual sensitivity to changes in vocal fold geometry and stiffness. *JASA* 129, 2529.