

Audibility of High Frequency Energy in Speech and Voice

Brian B. Monson

Speech, Language, and Hearing Sciences
University of Arizona

Sten Ternström

Speech, Music, and Hearing
Kunglia Tekniska Högskolan

Andrew J. Lotto

Speech, Language, and Hearing Sciences
University of Arizona

ACNS Conference

9 January 2009



Abstract

The human voice frequency spectrum above 5 kHz is often neglected in studies on the human voice. Perceptual studies involving this high-frequency range suggest that it may be of perceptual significance. Percepts possibly affected by this frequency range include speech intelligibility, talker discrimination, naturalness, and breathiness. To determine the validity of further research on this high-frequency range, a perceptual study was performed wherein listeners were required to discriminate between human voice stimuli that differed only in high-frequency content. High-frequency energy levels in human voice stimuli were manipulated to determine difference limens of discrimination, as well as thresholds of detection. All listeners showed sensitivity to the octave band centered at 8 kHz, while about half of listeners showed sensitivity to the octave band centered at 16 kHz. These results may impact any area where qualitative assessment of voice is necessary or important, including speech pathology, vocal pedagogy, voice synthesis, wide-band telephony, and the recording studio.



Project Design

- Stimuli
 - two male voices
 - 'loud' and 'soft' phonation, 30 cm
 - vowel /a/, constant F0, 500 ms
 - sound treated room, low noise
- Presented over Audio Pro type A4-14 full range speaker at 1 m

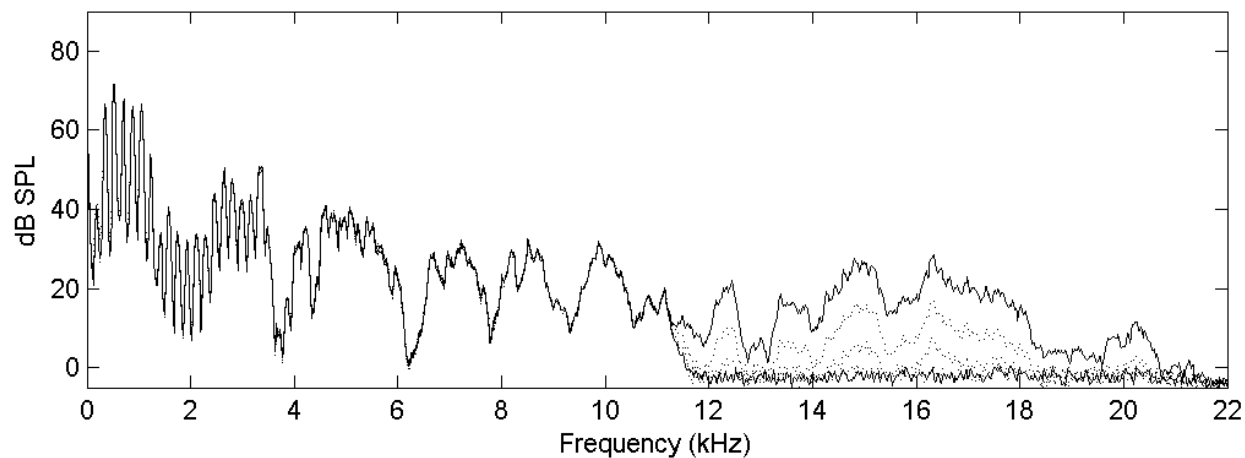
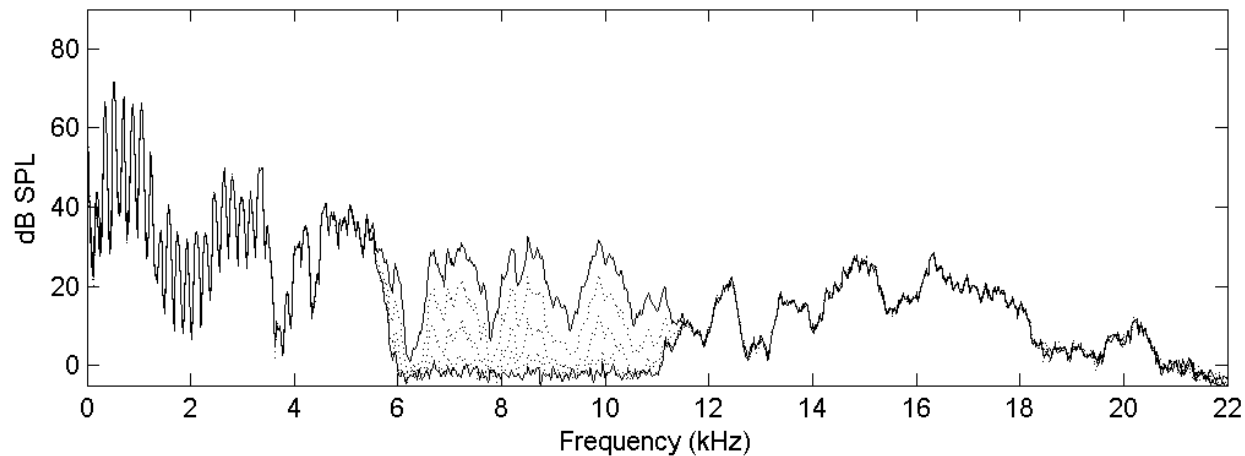


Project Design

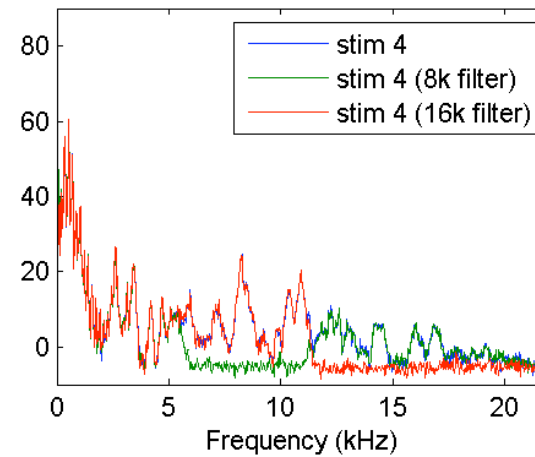
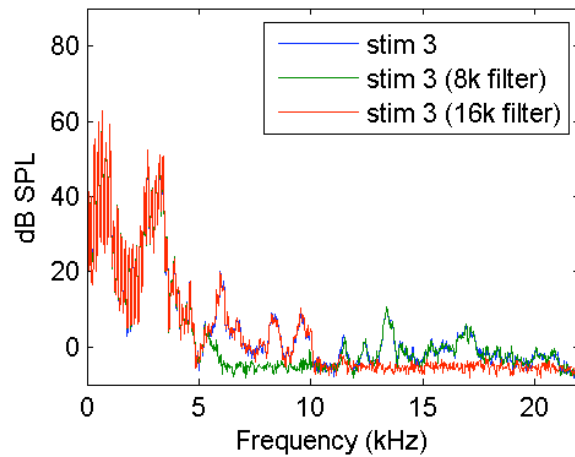
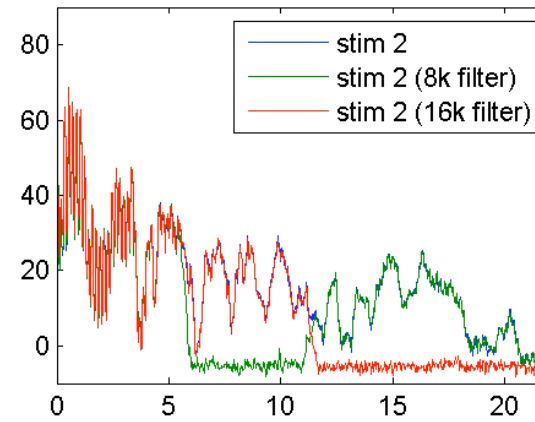
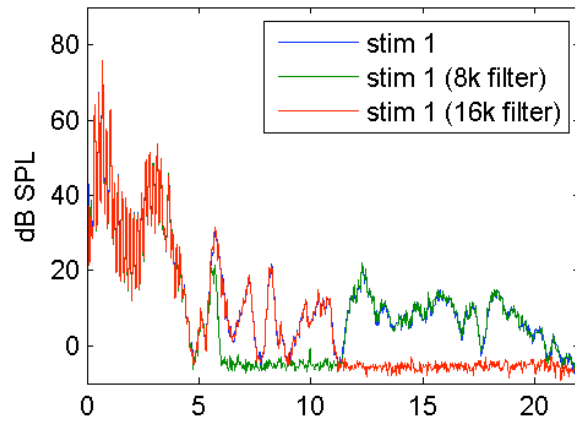
- Stimuli presented at realistic levels (based on initial recordings made at 30 cm)
- Stimuli
 - Filtered 8-kHz and 16-kHz octaves separately
 - Decreased octave level until listener could detect difference



Stimulus Example



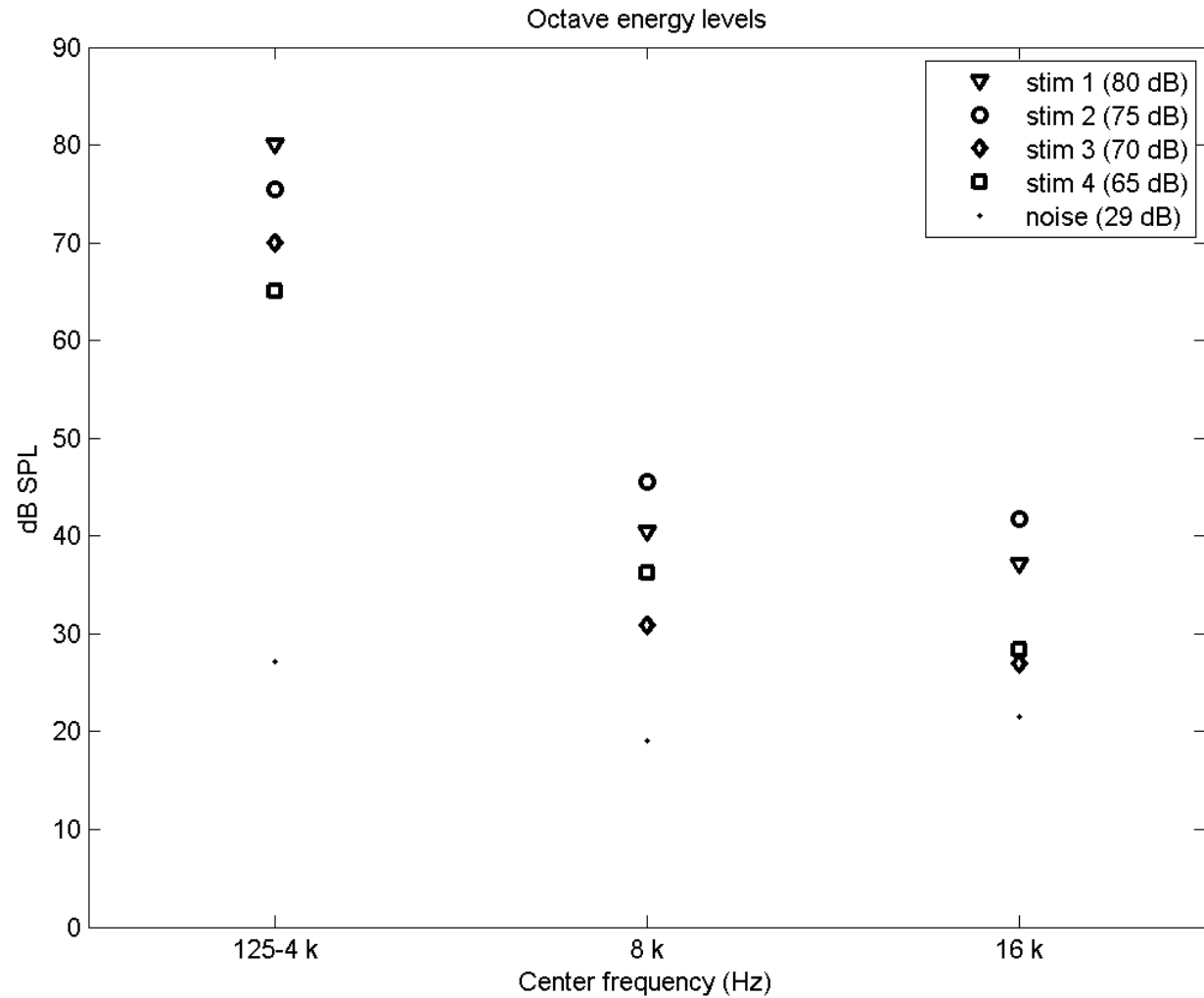
Stimuli



Stimuli HF Energy Levels

Talker 1:
stim 1 & 3

Talker 2:
stim 2 & 4

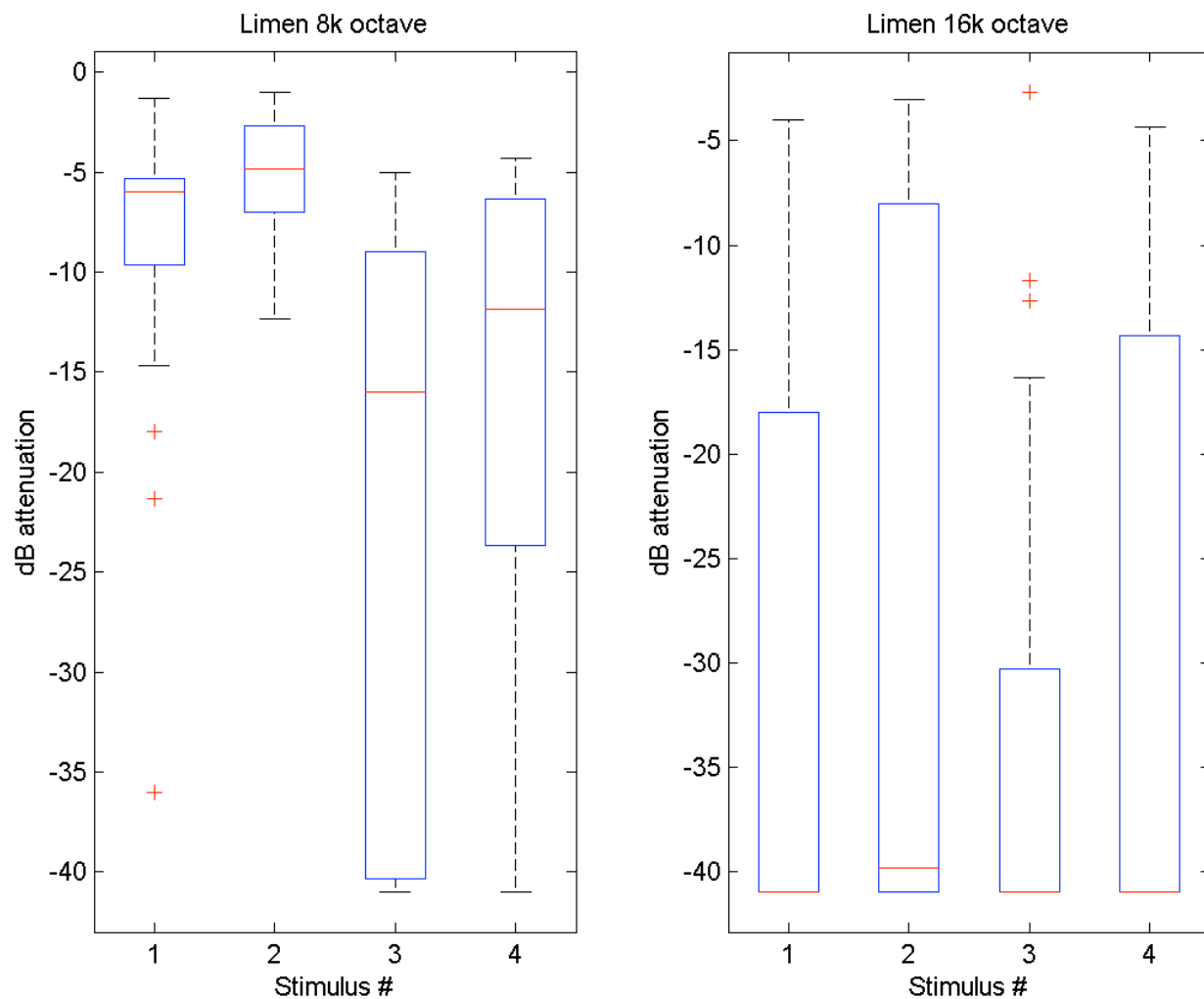


Project Design

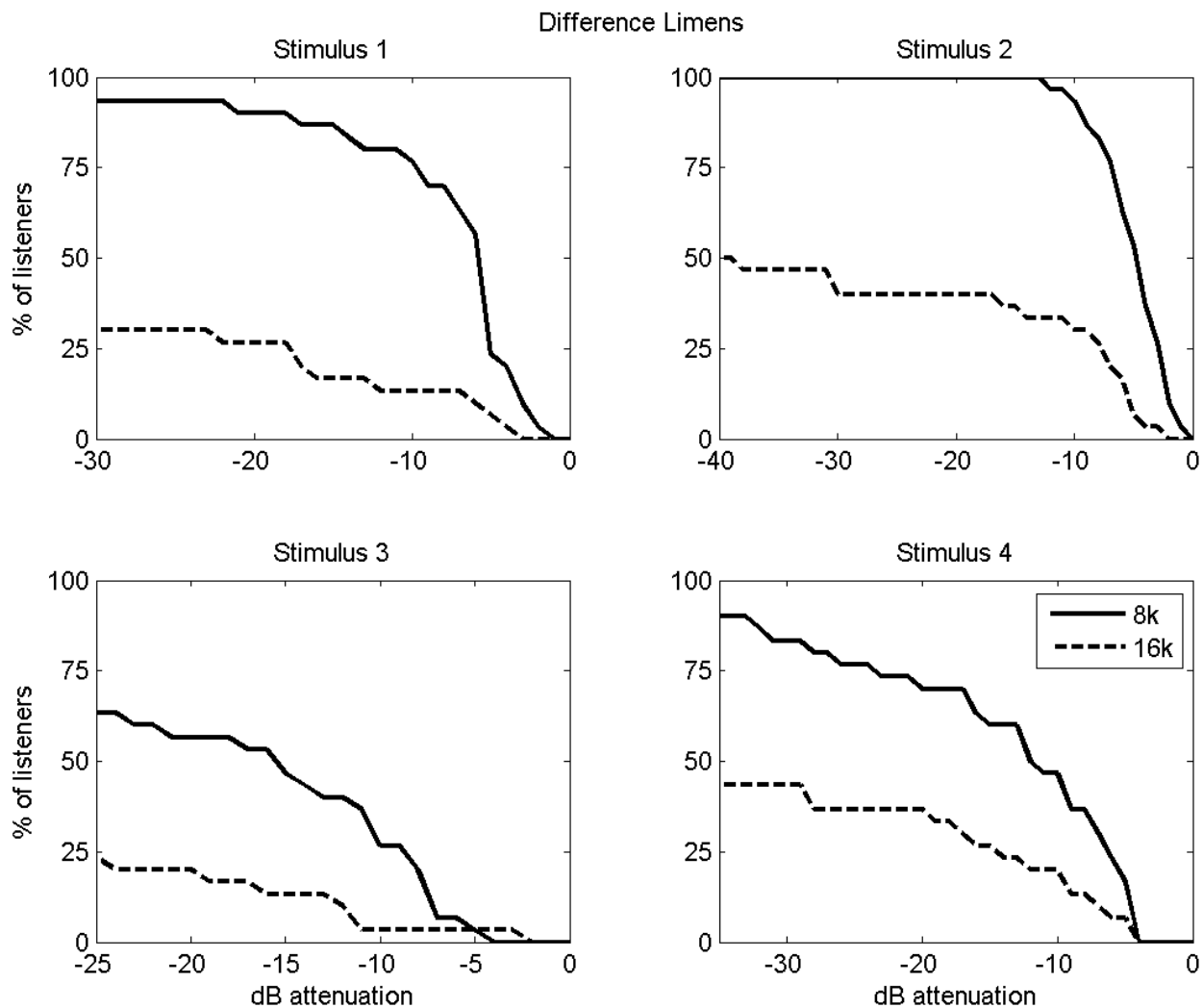
- Adaptive three-alternative forced-choice 1-up 2-down oddity task
 - Visual feedback ('CORRECT/INCORRECT')
- Listeners
 - 30 Listeners, 16 male, 14 female
 - Age range 20 - 60, mean 29.5
 - All listeners showed ability to detect at least some changes in HF energy



Results



Results



Conclusions

- Listeners can detect changes in 8-kHz octave
 - for phonation (no consonants)
 - appears to be dependent on HFE level, which is dependent on talker characteristics and phonation level
- Some listeners can detect changes in 16-kHz octave



Implications

- Characterization of perceptual information found within high-frequency energy appears to be worthwhile (germane to speech synthesis, vocal tract modeling, voice evaluation and management, CI, telephony, etc.)
- The Big Question: What is the perceptual relevance of high-frequency energy?

