Reliable Spectral Properties Elicit Contrast Effects in Perception of Noise-Vocoded Speech

PS-182

Christian Stilp

Department of Psychological and Brain Sciences

University of Louisville

INTRODUCTION

The auditory system is highly sensitive to stable aspects of the acoustic environment. This includes reliable spectral properties, or spectral peaks and shapes that are relatively stable or recurring across time. When a frequency region is reliable in preceding sounds but it changes in a subsequent target sound, the auditory system emphasizes this spectral difference, resulting in spectral contrast effects. For example, when preceding sounds have a reliable spectral peak in low-F_1 regions (sounds more "ih"-like), the following target vowel with a slightly higher F_1 sounds much higher by comparison, resulting in more "eh" responses, and *vice versa* (Ladefoged & Broadbent, 1957). Reliable spectral properties and spectral contrast effects have widespread impacts on speech perception by normal-hearing listeners (Ladefoged & Broadbent, 1957; Watkins, 1991; Holt, 2006; Stilp *et al*., accepted). However, despite their importance for normal-hearing listeners, how these phenomena influence speech perception by cochlear implant (CI) users is unclear. Here we used noise vocoding to investigate how CI users might respond to reliable spectral properties in spectrally degraded speech.

METHODS

Base Stimuli (from Stilp *et al.*, accepted)

Vowels: Natural productions of [ɪ] ("ih") and [ɛ] ("eh") (246 ms) were digitally edited to vary in mean F_1 from 415-565 Hz across a ten-step series.

Precursor Sentence: Vowels were preceded by a recording of the author saying "Please say what this vowel is" (2174 ms). Reliable spectral properties were added using FIR bandpass filters in MATLAB. Low F_1 (100-400 Hz) or high F_1 (550-850 Hz) regions were amplified in the precursor by +5 or +20 dB.

Noise Vocoding

Stimuli were noise-vocoded from 100-5000 Hz. This low-frequency edge can cause filter instability, but it is essential to preserve information from 100-400 Hz. To overcome this limitation, stimuli were spectrally rotated about 8000 Hz: 100 Hz in the original signal was transposed to 8000 Hz in the rotated signal, and 5000 Hz in the original signal (total stimulus bandwidth) was transposed to 3000 Hz in the rotated signal. Corner frequencies for channels from 100-5000 Hz were computed using Greenwood's (1990) formula then subtracted from 8100 Hz. Thus, acoustic frequencies and channel corner frequencies were both inverted but properly aligned as in typical vocoding. The spectrally rotated signal was noise-vocoded (4th-order Butterworth filters for channel analysis and synthesis; 2nd-order Butterworth filter with low-pass cutoff at 400 Hz for envelope extraction). The vocoded signal was spectrally rotated again about 8000 Hz to return all frequencies to their original positions.

Procedure

Listeners (28 native English speakers with normal hearing) heard stimuli over circumaural headphones and responded by clicking the mouse to indicate whether the target vowel sounded more like "ih" or "eh".

Analysis

Logistic regressions were fit to responses using generalized linear mixed-effects models in R. The final model took the following form:

response ~ vowel + filter + NCh + filter-by-NCh + vowel-by-NCh + (1 + vowel + filter + NCh | subject)

filter = filter frequency region (low F_1, high F_1)

NCh = number of spectral channels (6, 12, 24)

50% points (equal probability of "ih" and "eh" responses) were calculated for each regression function. Spectral contrast effect size was defined as the distance between 50% points, measured in stimulus steps along the abscissa.

EXPERIMENT 1: +20 dB PEAK (n = 14)

6 Channels | 12 Channels | 24 Channels | Full Spectrum (Stilp *et al*., accepted)

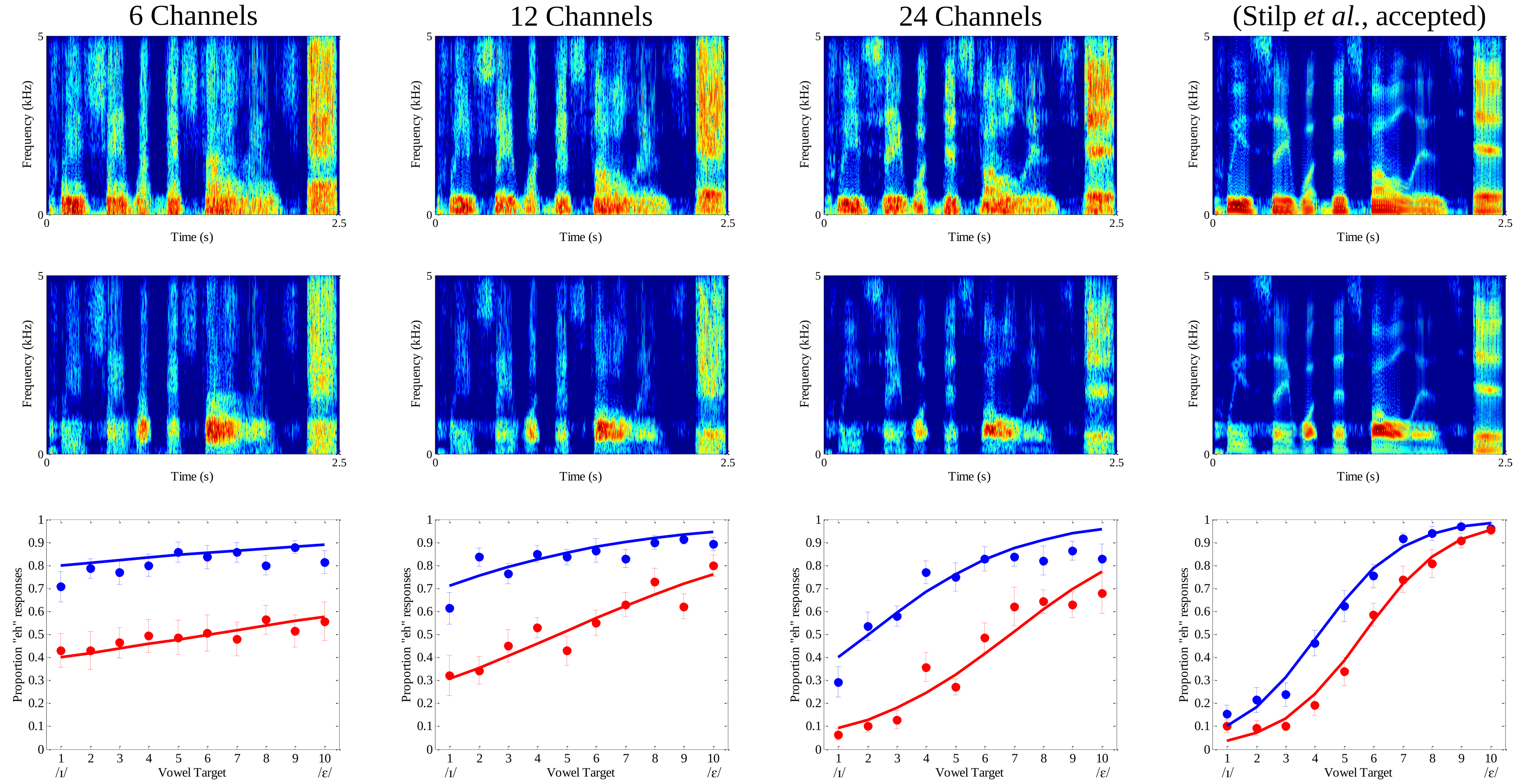
Low-F_1 peak

High-F_1 peak

Responses (Error bars = SEM)

Low-F_1 precursor

High-F_1 precursor



Contrast effect size: 22.2 steps | Contrast effect size: 7.8 steps | Contrast effect size: 4.9 steps | Contrast effect size: 1.5 steps

EXPERIMENT 2: +5 dB PEAK (n = 14)

6 Channels | 12 Channels | 24 Channels | Full Spectrum (Stilp *et al*., accepted)

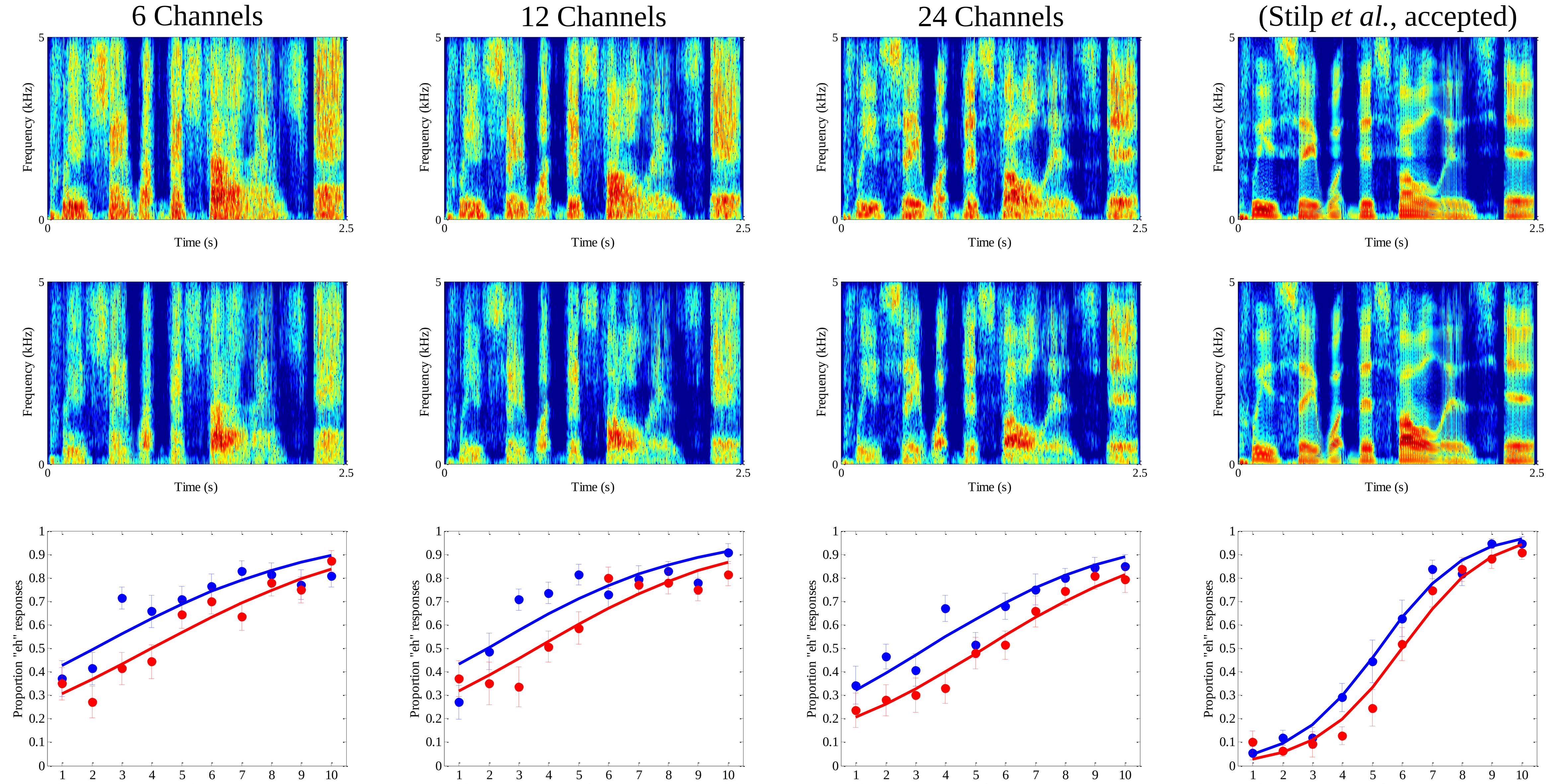
Low-F_1 peak

High-F_1 peak

Responses (Error bars = SEM)

Low-F_1 precursor

High-F_1 precursor



Contrast effect size: 1.9 steps | Contrast effect size: 1.6 steps | Contrast effect size: 1.9 steps | Contrast effect size: 0.8 steps

Upper cutoff frequencies for vocoder channels. Channel numbers are listed in the first three columns for each level of spectral resolution tested. Colored cutoff frequencies indicate which channels were filtered to add the reliable spectral property (low F_1: 100-400 Hz, high F_1: 550-850 Hz).

6 Channels	12 Channels	24 Channels	Channel Upper Cutoff (Hz)
1	1	1	135
		2	175
	2	3	219
		4	270
2	3	5	327
		6	392
	4	7	465
		8	549
3	5	9	643
		10	749
	6	11	869
		12	1006
4	7	13	1160
		14	1334
	8	15	1532
		16	1755
5	9	17	2008
		18	2294
	10	19	2618
		20	2984
6	11	21	3399
		22	3868
	12	23	4399
		24	5000

DISCUSSION

Auditory perception is sensitive to reliable spectral properties in spectrally degraded speech. Listeners responded "ih" (lower F_1) more often following precursors with reliable spectral properties consistent with [ɛ] (higher F_1) and *vice versa*, revealing spectral contrast effects in perception of noise-vocoded speech. This result was observed across all spectral resolutions tested (6-24 vocoder channels), suggesting widespread importance for perception of vocoded speech.

When preceding speech featured a +20 dB reliable spectral peak, spectral contrast effects grew as spectral resolution worsened. With only six spectral channels, listeners had difficulty accurately distinguishing target vowels, potentially due to long-term adaptation, masking, and/or frequency overlap between the reliable spectral peak and target vowel in such broad channels. However, responses were still heavily biased away from (contrastive with) the reliable spectral property. With 24 channels, contrast effects were still markedly larger than those observed in full-spectrum speech in Stilp *et al.* (accepted).

When preceding speech featured a +5 dB reliable spectral peak, spectral contrast effect size was relatively constant across all spectral resolutions. This is a different pattern from the +20 dB results, suggesting differential sensitivity to the magnitude of reliable spectral peaks. Similar to +20 dB results, contrast effects were still larger than those observed in full-spectrum speech with +5 dB reliable peaks (Stilp *et al*., accepted).

CI users experience auditory enhancement effects (Goupell & Mostardi, 2012; Wang *et al.*, 2012), revealing spectral changes are perceptually enhanced for these listeners. In addition, spectral changes are enhanced throughout the central auditory system (Scutt & Palmer, 1998; Nelson & Young, 2010), not only in the cochlea. Together with the present results, this bolsters the prediction that CI users are sensitive to reliable spectral properties and should also show spectral contrast effects.

REFERENCES

Goupell, M.J. & Mostardi, M.J. (2012) *J. Acoust. Soc. Am. 131(2)*, 1007-1010.
Greenwood, D.D. (1990) *J. Acoust. Soc. Am. 87(6)*, 2592-2605.
Holt, L.L. (2006) *J. Acoust. Soc. Am. 120(5)*, 2801-2817.
Ladefoged, P. & Broadbent, D.E. (1957) *J. Acoust. Soc. Am. 29(1)*, 98-104.
Nelson, P.C. & Young, E.D. (2010) *J. Neurosci. 30(19)*, 6577-6587.
Scutt, M.J. & Palmer, A.R. (1998) *ARO Abstracts, 381(A)*.
Stilp, C.E., Anderson, P.W., & Winn, M.B. (accepted) *J. Acoust. Soc. Am.*
Wang, N., Kreft, H., & Oxenham, A.J. (2012) *J. Acoust. Soc. Am. 131(6)*, EL421-EL426.
Watkins, A.J. (1991) *J. Acoust. Soc. Am. 90(6)*, 2942-2955.