Defining Essential Characteristics of Reliable Spectral Properties That Elicit Spectral Contrast Effects in Vowel Identification

1aPP24

UNIVERSITY OF LOUISVILLE

Paul W. Anderson and Christian E. Stilp

Department of Psychological and Brain Sciences, University of Louisville

INTRODUCTION

Perception enhances spectral differences between a preceding acoustic context with a reliable spectral property (*e.g.*, amplified frequency region, long-term spectral shape) and a subsequent target vowel sound, producing spectral contrast effects.

- Shifting F_1 frequencies downward in a preceding sentence (sounding more [ɪ]-like) increased the number of [ɛ] (high-F_1) responses and *vice versa* (Ladefoged & Broadbent, 1957)
- Sentence filtered to emphasize lower frequencies (spectrum of [ɪ]-minus-[ɛ]) produced more [ɛ] responses and *vice versa* (Watkins, 1991)

These contrast effects are very robust, influencing identification of vowels as well as stop consonants (Laing *et al*., 2012), fricatives (Watkins & Makin, 1996), and musical instruments (Stilp *et al*., 2010).

However, conditions conducive to contrast effects are still poorly understood:

- Perception calibrates to narrowband (100-Hz) spectral regularities that are reliable across both context and target (Kiefte & Kluender, 2008; Alexander & Kluender, 2010; Anderson & Stilp, 2014); are narrowband spectral peaks sufficient to elicit contrast effects as well?
- Do contrast effects require shifting formant ranges up or down to simulate different talkers, or is mere amplification of formant ranges sufficient?
- When are reliable spectral properties too weak (*e.g.*, insufficient amplification) to influence speech perception?
- Finally, what acoustic properties of these spectral regularities predict not only the presence of a contrast effect, but its magnitude as well?

Spectral contrast effects in vowel identification were explored using precursors filtered to have reliable narrowband (100 Hz) or broadband (300 Hz) peaks, or the difference between vowel spectral envelopes. Results reveal the efficacy of different spectral regularities in influencing speech perception.

METHODS

Precursor

- "Please say what this vowel is" spoken by CS (2174 ms)

Vowels

- Natural vowels linearly interpolated from [ɪ] to [ɛ] using PRAAT (246 ms)
- [ɪ] endpoint: $f0$ = 100 Hz, F_1 = 400→430 Hz, F_2 = 2000→1800 Hz
- [ɛ] endpoint: $f0$ = 100 Hz, F_1 = 580→550 Hz, F_2 = 1800→1700 Hz

Filters

- Narrowband (NB): 100-Hz bandwidth (250-350 or 550-650 Hz)
- Broadband (BB): 300-Hz bandwidth (100-400 Hz or 550-850 Hz)
- Spectral Envelope Difference (SED): spectral envelopes for [ɪ] and [ɛ] endpoints derived via FFT then subtracted from one another
- Filter gains set to +5 / +10 / +15 / +20 dB (NB, BB) or 25% / 50% / 75% / 100% of total filter power (SED)

Participants

- All native English speakers with normal hearing
- n = 13 (NB20, BB20, SED100%), n = 14 (NB15, NB10, NB5), n = 11 (BB15, BB10, BB5), n = 11 (SED75%, SED50%, SED25%)

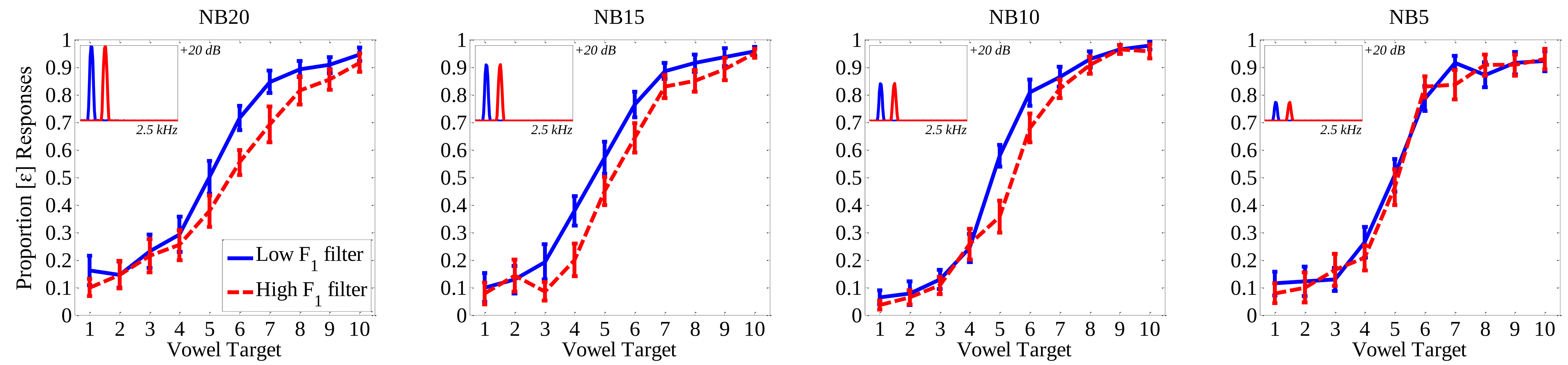
Procedure

- All precursor – vowel pairs presented diotically at 70 dB SPL via circumaural headphones in sound-isolating booths
- Filter types were blocked and tested in random orders

RESULTS

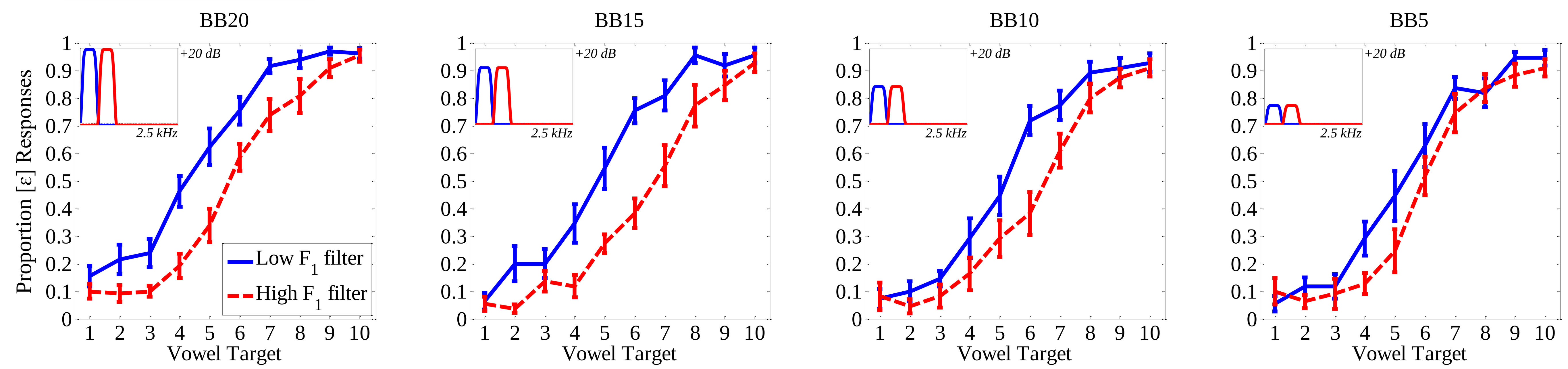
- Logistic regressions were fit to each listener's identification curves. Midpoints were calculated from these regression functions.
- Shift in midpoints across filtering conditions (*i.e.*, translation along the abscissa) measures the magnitude of the contrast effect.
- Midpoint shifts were analyzed using paired-sample t-tests (Bonferroni correction for multiple analyses within each participant group; α = .05 / 3 = .0167).
- Mean responses are plotted below, with filters that processed the precursor depicted in inset. * indicates statistically significant midpoint shifts; error bars depict ±1 SEM.

NARROWBAND FILTERS



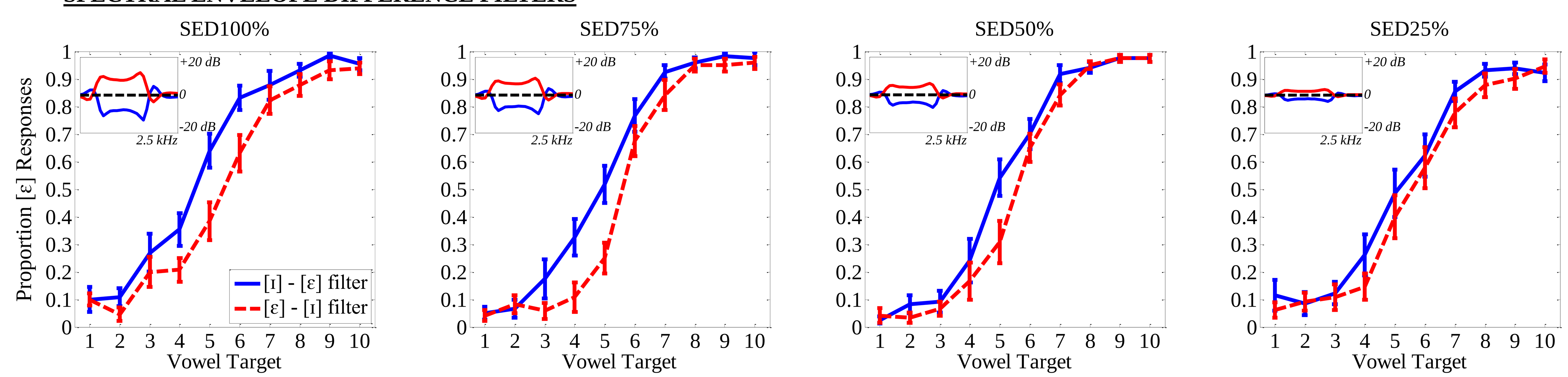
NB20	NB15	NB10	NB5
Mean midpoint shift = 1.07 steps*	Mean midpoint shift = 0.86 steps*	Mean midpoint shift = 0.53 steps*	Mean midpoint shift = 0.17 steps
t_{12} = 4.44, p < .001	t_{13} = 5.99, p < .0001	t_{13} = 4.44, p < .01	t_{13} = 0.81, p = .43

BROADBAND FILTERS



BB20	BB15	BB10	BB5
Mean midpoint shift = 1.67 steps*	Mean midpoint shift = 1.90 steps*	Mean midpoint shift = 1.10 steps*	Mean midpoint shift = 0.77 steps*
t_{12} = 6.35, p < .0001	t_{10} = 5.85, p < .001	t_{10} = 6.53, p < .0001	t_{10} = 4.67, p < .001

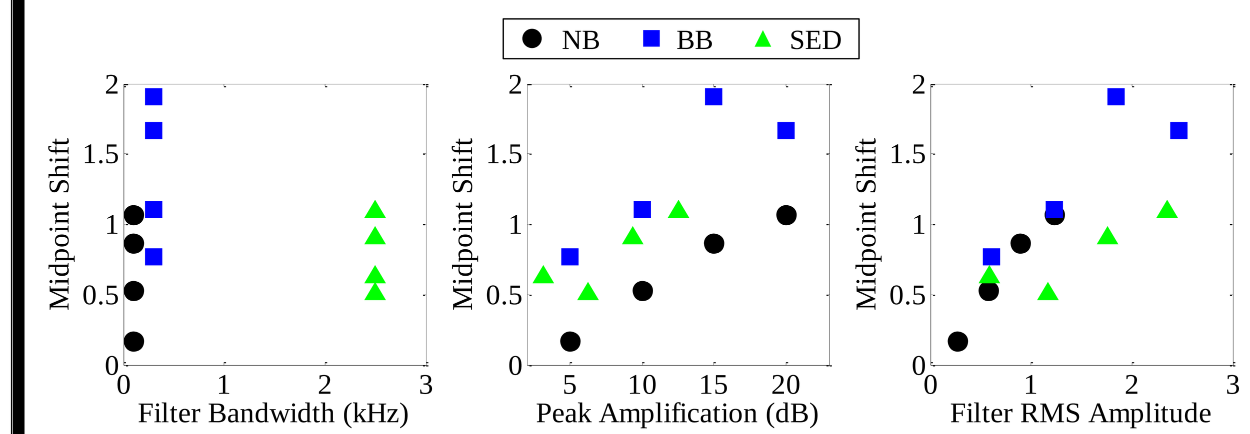
SPECTRAL ENVELOPE DIFFERENCE FILTERS



SED100%	SED75%	SED50%	SED25%
Mean midpoint shift = 1.09 steps*	Mean midpoint shift = 0.91 steps*	Mean midpoint shift = 0.50 steps	Mean midpoint shift = 0.63 steps*
t_{12} = 5.59, p < .001	t_{10} = 4.82, p < .001	t_{10} = 2.86, p = .017	t_{10} = 3.17, p < .01

ACOUSTIC PREDICTORS

Vowel identification shifted following most but not all spectral regularities in the filtered precursor. What predicts these results?



Left: Filter bandwidth is a poor predictor of contrast effect magnitude (measured using mean midpoint shift) (r = -0.18, p = 0.57)

Middle: Greater peak amplification of key frequency regions correlates well with larger contrast effects (r = 0.72, p < .01)

Right: Total filter power (measured by RMS amplitude) also predicts results (r = 0.78, p < .005)

Multiple regression: Total filter power reliably predicts results (β = 0.91, t = 2.85, p < .025) but peak amplification does not (β = -0.03, t = -0.98, p = .92)

CONCLUSIONS

- Spectral contrast effects in vowel identification are demonstrated for very modest peaks in the precursor spectrum (as little as 5 dB) or very narrowband frequency regions (100-Hz wide).
  - Reveals acute sensitivity to a very broad range of reliable spectral properties of a listening context
  - Contrast effects may be more pervasive in speech perception than previously thought
- Total filter power is the most complete predictor of contrast effect magnitude.
  - Explains tradeoff between bandwidth and peak amplification for NB and BB regularities (NB20 results ≈ BB10 results; NB10 ≈ BB5)
- Contrast effects are observed for spectral regularities in speech that are not directly derived from speech or vocal tracts (NB, BB).
  - Replicates classic findings by Ladefoged and Broadbent (1957); consistent with general auditory account of speech perception (*e.g.*, Laing *et al*., 2012)
- Generalizability of results across speech and nonspeech targets, speech and nonspeech precursors, and a wide variety of reliable spectral properties reveals optimizing sensitivity to change is a fundamental operating characteristic of the auditory system.

REFERENCES

Alexander & Kluender (2010) *JASA*
Anderson & Stilp (2014) *ARO*
Kiefte & Kluender (2008) *JASA*
Ladefoged & Broadbent (1957) *JASA*
Laing, Liu, Lotto, & Holt (2012) *Frontiers in Psych.*
Stilp, Alexander, Kiefte, & Kluender (2010) *JASA*
Watkins (1991) *JASA*
Watkins & Makin (1996) *JASA*