Auditory enhancement and spectral contrast effects in speech perception[a)]

Christian E. Stilp[b)]
317 Life Sciences Building, University of Louisville, Louisville, Kentucky 40292, USA

(Received 11 January 2019; revised 10 July 2019; accepted 11 July 2019; published online 30 August 2019)

The auditory system is remarkably sensitive to changes in the acoustic environment. This is exemplified by two classic effects of preceding spectral context on perception. In auditory enhancement effects (EEs), the absence and subsequent insertion of a frequency component increases its salience. In spectral contrast effects (SCEs), spectral differences between earlier and later (target) sounds are perceptually magnified, biasing target sound categorization. These effects have been suggested to be related, but have largely been studied separately. Here, EEs and SCEs are demonstrated using the same speech materials. In Experiment 1, listeners categorized vowels (/ɪ/-/ɛ/) or consonants (/d/-/g/) following a sentence processed by a bandpass or bandstop filter (vowel tasks: 100–400 or 550–850 Hz; consonant tasks: 1700–2700 or 2700–3700 Hz). Bandpass filtering produced SCEs and bandstop filtering produced EEs, with effect magnitudes significantly correlated at the individual differences level. In Experiment 2, context sentences were processed by variable-depth notch filters in these frequency regions (−5 to −20 dB). EE magnitudes increased at larger notch depths, growing linearly in consonant categorization. This parallels previous research where SCEs increased linearly for larger spectral peaks in the context sentence. These results link EEs and SCEs, as both shape speech categorization in orderly ways. © *2019 Acoustical Society of America.*
https://doi.org/10.1121/1.5120181

[VMR] Pages: 1503–1517

I. INTRODUCTION

All perception takes place in context. In auditory perception, two classic findings typify the highly influential role of surrounding acoustic context. First, the salience and/or detectability of particular frequencies is shaped through auditory enhancement effects (EEs) (Schouten, 1940; Viemeister, 1980; Viemeister and Bacon, 1982; Thibodeau, 1991; Byrne *et al.*, 2011; Erviti *et al.*, 2011; Carcagno *et al.*, 2012; Feng and Oxenham, 2015, 2018a; Kreft *et al.*, 2018). For example, in a simultaneous masking paradigm, the target frequency is embedded in a multitone complex. Detection thresholds are measured for this masker-plus-target complex alone and when following an adaptor stimulus with the same spectrum minus the target frequency. Detection of the target frequency is improved in this latter case. In a forward masking paradigm, the masker stimulus precedes an isolated target tone. Detection thresholds are better in this case compared to when an adaptor stimulus with the same spectrum as the masker minus energy at the target frequency precedes (and enhances) the masker. This general experimental paradigm has been expanded broadly, from enhancing frequencies corresponding to formant peaks and producing vowel percepts (Summerfield *et al.*, 1984; Summerfield *et al.*, 1987), binaural centering tasks (Byrne *et al.*, 2011), pitch salience judgments for the enhanced component (Byrne *et al.*, 2013), and judgments regarding the presence/absence or pitch movement of components (Erviti *et al.*, 2011; Carcagno *et al.*, 2012; Feng and Oxenham, 2015).

Second, perception of complex sounds is also influenced by spectral contrast effects (SCEs) (Ladefoged and Broadbent, 1957; Lotto and Kluender, 1998; Watkins, 1991; Holt, 2005, 2006; Stilp *et al.*, 2010; Stilp *et al.*, 2015; Kingston *et al.*, 2014; Stilp and Assgari, 2017, 2018, 2019; Feng and Oxenham, 2018b). A typical SCE paradigm measures categorization of a continuum of target sounds following contexts with different spectral properties. In a seminal paper, Ladefoged and Broadbent (1957) demonstrated that the spectrum of a preceding sentence context ("Please say what vowel this is") influenced categorization of subsequent vowel targets. When the first formant frequencies (F_1) of the context sentence were shifted up to higher frequencies, listeners perceived the subsequent target vowel as /ɪ/ (low F_1) more often. When the first formant frequencies of the context sentence were shifted down, listeners perceived the target as /ɛ/ (high F_1) more often. Decades of research have demonstrated the flexibility and robustness of this effect: from short-term (a single sound or syllable; e.g., Lotto and Kluender, 1998) to long-term context effects (a series of sounds or a sentence; e.g., Ladefoged and Broadbent, 1957); across a wide range of speech sounds (see Stilp *et al.*, 2015 for review); from nonspeech contexts (*e.g.*, Watkins, 1991; Holt, 2005) to nonspeech targets (Stilp *et al.*, 2010; Kingston *et al.*, 2014; Frazier *et al.*, 2019); and from highly controlled filtered contexts to less controlled unfiltered contexts (Stilp and Assgari, 2019).

EEs and SCEs exhibit several broad similarities. First, both demonstrate enhanced processing of spectral differences

[a)]Portions of these results were presented at the 176th (Victoria, British Columbia, Canada) and 177th (Louisville, KY, USA) Meetings of the Acoustical Society of America.
[b)]Electronic mail: christian.stilp@louisville.edu

over time. In EEs, some frequency components are present throughout the entire trial, while other components (specifically, the target frequency) are often only introduced later in the trial. Audibility of these introduced frequencies becomes perceptually enhanced. In SCEs, changes in spectral composition across the context and target are perceptually magnified, biasing categorization of the subsequent target sound. As such, these context effects have been suggested to be related to processes that increase the auditory system's sensitivity to changes in the acoustic environment (Feng and Oxenham, 2015, 2018c). Second, the mechanisms underlying each effect are thought to reside in relatively low-level neural processing. In simple neural adaptation, neurons responding to frequencies in earlier (context) sounds would be adapted and thus less responsive upon introduction of later (target) sounds. Neurons encoding other frequencies would be unadapted/less adapted and thus relatively more responsive upon introduction of the target sound. Simple neural adaptation has been suggested to produce SCEs (Delgutte, 1996; Delgutte *et al.*, 1996; Holt *et al.*, 2000; Holt and Lotto, 2002; Stilp and Assgari, 2018; but see Summerfield *et al.*, 1984; Summerfield *et al.*, 1987 for arguments that effects thought to be produced by adaptation might be instead produced by adaptation of suppression/inhibition). Simple neural adaptation was initially thought to underlie EEs as well (Viemeister, 1980). Subsequent studies proposed that these effects were instead produced by adaptation of suppression/inhibition, where the inhibitory influence of activated neurons over those encoding neighboring frequencies adapts over time. This results in neural responses to the suppressed/inhibited frequencies being more pronounced than they were initially (e.g., Viemeister and Bacon, 1982; Summerfield *et al.*, 1984; Nelson and Young, 2010; Byrne *et al.*, 2011; Carcagno *et al.*, 2012; Wang and Oxenham, 2016). Finally, some results suggest that these effects may be bidirectional in time (to a degree). The typical testing paradigm for each context effect presents a listening context before the target sound (i.e., forwards context effect), but instances of both EEs (Kidd and Wright, 1994; Byrne *et al.*, 2013) and SCEs (Winn *et al.*, 2013; Sjerps *et al.*, 2018) have been reported when the target precedes the context (i.e., backwards context effect, but with short or no interstimulus intervals and often smaller magnitudes than the forwards effect). These similarities have led researchers to suggest that EEs and SCEs are related to each other (Holt and Lotto, 2002; Kluender *et al.*, 2003; Feng and Oxenham, 2018a).

Two points potentially temper the closeness of the relationship between EEs and SCEs. First, EEs and SCEs are measured and quantified quite differently. EEs tend to be quantified as changes in masked thresholds (often in dB) for enhanced versus unenhanced conditions. Other reports offer alternative quantifications of EEs, such as changes in detection (Erviti *et al.*, 2011; Carcagno *et al.*, 2012), pitch salience (Byrne *et al.*, 2013), or vowel recognition accuracy (Summerfield *et al.*, 1984; Summerfield *et al.*, 1987). Conversely, SCEs tend to be quantified as shifts in target categorization (whether quantified as changes in the rate of a particular response, shifts in the category boundary, or some other related parameter). Second, the stimuli used to measure these effects tend to be very different. Most EE paradigms have employed tone complexes or band-limited noise in order to have tight experimental control over their spectral and temporal characteristics. On the other hand, SCEs have largely been measured in speech categorization tasks, often following speech contexts. This trend is not exclusive, as noise or pure-tone contexts can bias speech categorization (Watkins, 1991; Lotto and Kluender, 1998; Holt, 2005), and speech or nonspeech contexts can bias categorization of tones or musical instruments (Stilp *et al.*, 2010; Kingston *et al.*, 2014; Frazier *et al.*, 2019).

Two studies compared contrastive and enhancement-based processes in speech perception. Coady *et al.* (2003) measured /ba/-/da/ categorization following different brief (100–300 ms) harmonic complex contexts. Some contexts resembled vowel spectra, with harmonics located at low-F_2 (720–960 Hz, appropriate for /o/) or high-F_2 (1800–2040 Hz, appropriate for /e/) frequencies to encourage perception of /da/ or /ba/, respectively, through SCEs. Other contexts were complementary to these spectra, replacing harmonics with silence and silence with harmonics, forming spectra with notches at vowel formant frequencies. Vowel-type spectra had contrastive effects on subsequent consonant categorization (consistent with SCEs) and notch-type spectra had assimilative effects on consonant categorization (consistent with EEs). Later, Holt (2006) tested a version of this paradigm using longer (2100 ms) contexts consisting of pure tones. Contexts with a higher spectral mean encouraged low-F_3-onset /ga/ responses, and contexts with a lower spectral mean encouraged high-F_3-onset /da/ responses. Complementary versions of these spectra, where tones were replaced by silence and silence was replaced by wideband noise (i.e., noise with spectrotemporal notches where tones used to be), had complementary effects on performance: contexts with high-frequency notches encouraged more high-F_3-onset /da/ responses, and contexts with lower-frequency notches encouraged more low-F_3-onset /ga/ responses.

These studies reported contrastive (SCE) and assimilative (EE) effects of preceding context on speech categorization, but had several limitations. First, each study reported contrastive and assimilative effects within the same subject group, but did not consider individual differences in effect magnitudes. If a given listener exhibited a relatively large SCE, did s/he also exhibit a relatively large EE? Analyses of such relationships would shed considerable light on the extent to which these effects are related. Second, EE-type effects were reported but their relative magnitudes were not considered. Recent work has demonstrated that SCE magnitudes vary continuously as a function of the spectral difference between context and target stimuli (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2017, 2019; Frazier *et al.*, 2019). EE magnitudes vary as functions of adaptor duration, interstimulus interval duration, and spectral notch width (Viemeister, 1980; Summerfield *et al.*, 1984; Viemeister *et* al., 2013; Feng and Oxenham, 2015; Kreft *et al.*, 2018), but whether these effects vary linearly in a similar manner to SCEs is unclear. Third, a very wide range of context stimuli have produced SCEs, ranging from high (speech, music) to low (tones, noise) in ecological validity (and high to low in spectrotemporal variability, respectively). However, contexts in EE experiments tend to be low

in ecological validity (and low in spectrotemporal variability: multitone complexes, notched noise, etc.). Thus, it is unclear whether EEs are also observed when using naturalistic (and highly acoustically variable) stimuli or only using more carefully controlled stimuli. SCEs have been argued to be pervasive in auditory perception (Stilp *et al.*, 2015; Stilp and Assgari, 2019), so if EEs are related to SCEs, they should be observable using more naturalistic stimuli with greater spectrotemporal variability as well.

The present experiments had three goals: (1) to study SCEs and EEs in speech perception using naturalistic stimuli (sentence contexts and phoneme targets); (2) to evaluate individual differences in context effect magnitudes; and (3) to test whether EE magnitudes vary continuously as a function of the spectral difference between context and target. In Experiment 1, listeners categorized vowels (/ɪ/-/ɛ/) or consonants (/d/-/g/) following a sentence context processed by a bandpass or bandstop filter. Bandpass filtering in low (for vowel categorization tasks, 100–400 Hz; for consonant categorization tasks: 1700–2700 Hz) or high (for vowel tasks, 550–850 Hz; for consonant tasks: 2700–3700 Hz) frequency regions was predicted to produce SCEs, as energy in the passband would be prominent compared to the rest of the (removed) spectrum. Bandstop filtering to remove energy from these same frequency regions was predicted to produce EEs, as the target sounds' energy at the context bandstop frequencies would be perceptually prominent. Calculation of context effects by subject permitted evaluation of individual differences across effect type (SCE, EE) and target type (vowels, consonants). Experiment 2 processed context sentences with variable-gain spectral notch filters (from −5 to −20 dB) to test whether EE magnitudes varied continuously as a function of spectral differences between context and target, as SCEs do (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2017, 2019; Frazier *et al.*, 2019). Experiment 2a tested this question in consonant categorization and Experiment 2b tested this question in vowel categorization.

II. EXPERIMENT 1

A. Methods

1. Participants

Twenty-one undergraduate students participated in Experiment 1 in exchange for course credit. All self-reported being native English speakers with no known hearing impairments.

2. Stimuli

a. Contexts. Experiment 1 employed two different context stimuli. In the vowel categorization task, the context stimulus was a recording of the author saying "Please say what this vowel is" (2174 ms). This stimulus has successfully biased vowel categorization through SCEs in previous research (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2018; 2019). In the consonant categorization task, the context stimulus was a recording of a male talker saying "Correct execution of my instructions is crucial" (2200 ms) from the TIMIT database (Garofolo *et al.*, 1990). This stimulus has successfully biased consonant categorization through SCEs in previous research (Stilp and Assgari, 2017, 2018). Spectral energy was approximately equal (within 1 dB) across the key frequency regions for each task [100–400 Hz (low F_1) and 550–850 Hz (high F_1) for the vowel task; 1700–2700 Hz (low F_3) and 2700–3700 Hz (high F_3) for the consonant task].

Each context sentence was processed by bandpass or bandstop filters applied to each of these frequency regions separately. Filters were zero-phase finite impulse response with 1000 coefficients and 1-Hz transition regions. This created four renditions of each sentence (vowel task: low-F_1 passband, high-F_1 passband, low-F_1 stopband, high-F_1 stopband; consonant task: low-F_3 passband, high-F_3 passband, low-F_3 stopband, high-F_3 stopband; see Fig. 1). All filters were created using the fir2 function in MATLAB (MathWorks, Inc., Natick, MA). Following filtering, all contexts were low-pass filtered at a cutoff frequency of 5000 Hz.

b. Targets. Vowels: Target vowels were the same /ɪ/-to-/ɛ/ continuum previously shown to be biased by SCEs (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp, 2017; Stilp and Assgari, 2018, 2019). For a detailed description of the generation procedures, see Winn and Litovsky (2015). Briefly, tokens of /ɪ/ and /ɛ/ were recorded by the author. Formant contours from each token were extracted using Praat (Boersma and Weenink, 2017). In the /ɪ/ endpoint, F_1 linearly increased from 400 to 430 Hz and F_2 linearly decreased from 2000 to 1800 Hz. In the /ɛ/ endpoint, F_1 linearly decreased from 580 to 550 Hz and F_2 linearly decreased from 1800 to 1700 Hz. These F_1 trajectories were linearly interpolated to create a ten-step continuum of formant tracks; linear interpolations also were performed for F_2 trajectories. The spectrum of the /ɪ/ endpoint was estimated using Burg's linear predictive coding procedure in Praat; it was then used to inverse filter the token in order to elicit the residual voice source. Formant tracks were used to filter this voice source, producing the ten-step continuum of vowel tokens. Energy above 2500 Hz was replaced with the energy high-pass-filtered from the original /ɪ/ token for all vowels. Final vowel stimuli were 246 ms in duration with fundamental frequency set to 100 Hz throughout the vowel.

Consonants: Target consonants were the same /da/-to-/ga/ continuum previously shown to be biased by SCEs (Stilp and Assgari, 2017, 2018). These ten morphed natural tokens were taken from Stephens and Holt (2011). F_3 onset frequencies varied from 2703 Hz (/da/ endpoint) to 2338 Hz (/ga/ endpoint) before converging at/near 2614 Hz for the following /a/. The duration of the consonant transition was 63 ms, and total syllable duration was 365 ms.

All context and target sounds were set to equal root-mean-square (rms) amplitudes. Trial sequences were then created by concatenating one target to a filtered context sentence with a 50-ms silent interstimulus interval. Finally, all stimuli were resampled at 44.1 kHz for presentation.

3. Procedure

After obtaining informed consent, each participant was led into a sound-attenuating booth (Acoustic Systems, Inc.,

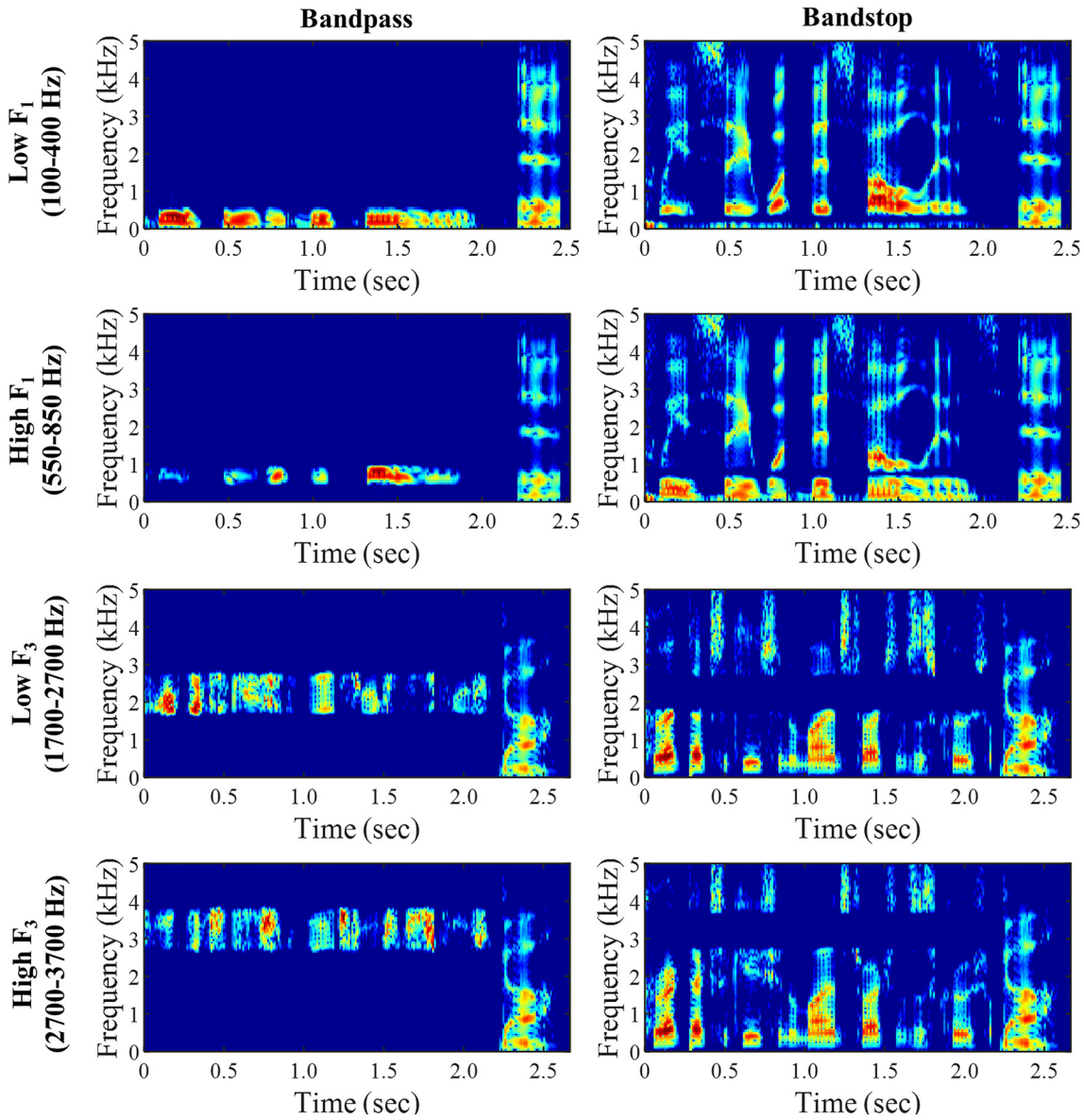

FIG. 1. (Color online) Sample trials from Experiment 1. Each row depicts one of the frequency regions manipulated in the context sentences. The top two rows depict stimuli from the vowel categorization tasks; the bottom two rows depict stimuli from the consonant categorization tasks. Columns depict the two methods of context sentence filtering: bandpass (left, predicted to produce SCEs) and bandstop (right, predicted to produce EEs).

Austin, TX). The participant sat at a small table on top of which was a computer screen, mouse, and keyboard. All sounds were D/A converted by RME HDSPe AIO sound cards (Audio AG, Haimhausen, Germany) on a personal computer and passed through a programmable attenuator (TDT PA4, Tucker-Davis Technologies, Alachua, FL) and headphone buffer (TDT HB6) before being presented diotically at a mean level of 70 dB sound pressure level (SPL) over circumaural headphones (Beyerdynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY). A custom MATLAB script guided participants through the experiment. After each trial, participants clicked the mouse to indicate what the target phoneme sounded like. In vowel tasks, they clicked on buttons to respond "ih (as in 'bit')" or "eh (as in 'bet')." In consonant tasks, they clicked on buttons to respond "da" or "ga."

First, the participant completed two practice sessions. The first practice session consisted of 20 trials presenting a context sentence from the AzBio corpus (Spahr *et al.*, 2012) followed by one of the two endpoints from the vowel continuum. Feedback was provided on practice trials. Listeners were required to categorize vowels with at least 80% accuracy. If they failed to meet this criterion, they were allowed to repeat this practice session up to two more times. If participants were still unable to categorize vowels with 80% accuracy after the third practice session, they were not allowed to participate in the main experiment. The second practice session was similarly structured but tested AzBio sentences preceding endpoints of the consonant continuum. Listeners were again required to categorize consonants with at least 80% accuracy, and were allowed to repeat this practice session up to two more times as needed. If listeners passed the vowel practice session but failed to pass the consonant practice session, they were not allowed to participate in the main experiment.

Second, listeners proceeded to the main experiment, which consisted of four blocks: F_1 passband contexts, F_1 stopband contexts, F_3 passband contexts, F_3 stopband contexts. Block order was counterbalanced across participants. Each block consisted of 160 trials (2 contexts $\times$ 10 targets $\times$ 8 repetitions), which were presented in random orders. The experiment was self-paced, allowing the participants opportunities to take breaks between each block. No feedback was provided. The entire experiment lasted approximately 40 min.

B. Results

One listener failed to pass the vowel practice session, leaving 20 listeners to participate in the main experiment. A second performance criterion was implemented where participants were required to maintain 80% accuracy on endpoint stimuli across related (vowels or consonants) blocks of the main experiment. Given the stated aim of analyzing individual differences across the four blocks, comparisons would be hampered if a participant met the criterion in only some experimental blocks. Five listeners failed to meet this criterion (all performing at <80% correct for vowel endpoints), so their data were removed from subsequent analyses.

1. Group-level analyses

Responses from the remaining 15 participants were analyzed using a mixed-effects logistic model in R (R Development Core Team, 2016) using the lme4 package (Bates *et al.*, 2014). The dependent variable was modeled as binary [lower-frequency responses ("ih," "ga") coded as 0 and higher-frequency responses ("eh," "da") coded as 1]. Fixed effects in the model included Target (coded as a continuous variable from 1 to 10 then mean-centered),[1] Frequency (contrast coded with the higher-frequency region coded as −0.5 and the lower-frequency region coded as +0.5), Phonetic Segment (contrast coded with vowel stimuli coded as −0.5 and consonant stimuli coded as +0.5), Filter Type (contrast coded with bandpass filtering coded as −0.5 and bandstop filtering coded as +0.5),[2] and the interactions between fixed effects. Random slopes were included for each fixed main effect, and a random intercept of participant was also included.

Results from the mixed-effects model are listed in Table I and illustrated in Fig. 2.[3] Results directly bearing on SCEs and EEs will be discussed first, followed by other significant influences on listeners' responses. These context effects cannot be evaluated when averaging across either Frequency (low or high frequency manipulation) or Filter Type (bandpass or bandstop). Bandpass filtering (predicted to produce SCEs) is predicted to increase responses at frequencies away from the spectral content of the preceding sentence; bandstop filtering (predicted to produce EEs) is predicted to increase responses at frequencies of the spectral notch in the preceding sentence. Examining Frequency or Filter Type alone collapses across opposing influences on listeners' responses, producing null results (the nonsignificant effects of Frequency or Filter Type by themselves in Table I). When these factors are evaluated together, their opposing influences are evident (significant Frequency by Filter Type interaction). Changing the filter from high frequencies to low frequencies increased high-frequency responses under bandpass filtering (via SCEs) but decreased high-frequency responses under bandstop filtering (via EEs). This was observed whether collapsing across consonant and vowel tasks (the significant two-way Frequency by Filter Type interaction) or extending the analysis to include each task (the significant three-way interaction between Frequency, Phonetic Segment, and Filter Type). Interactions between Phonetic Segment and only one of these factors averaged across critical manipulations responsible for producing shifts in opposite directions (nonsignificant Frequency by Phonetic Segment interaction collapsed across Filter Type; nonsignificant Phonetic Segment by Filter Type interaction collapsed

TABLE I. Mixed-effects model analysis of responses in Experiment 1. Columns indicate each effect, the model coefficient (β), standard error of the mean (SE), and corresponding Z statistic and p-value for that term. See main text for description of the model architecture. Contrast-coded factors are followed by the level associated with the −0.5 contrast in parentheses.

Effect	β	SE	z	p
Intercept	0.154	0.148	1.036	0.300
Target	1.207	0.062	19.398	<0.001
Frequency (high)	−0.241	0.145	−1.665	0.096
Phonetic Segment (vowel)	−0.914	0.331	−2.758	0.006
Filter Type (passband)	0.008	0.172	0.046	0.964
Target × Frequency	−0.197	0.046	−4.300	<0.001
Target × Phonetic Segment	0.333	0.048	6.998	<0.001
Frequency × Phonetic Segment	0.006	0.158	0.038	0.970
Target × Filter Type	0.089	0.046	1.943	0.052
Frequency × Filter Type	−3.158	0.160	−19.745	<0.001
Phonetic Segment × Filter Type	0.236	0.158	1.496	0.135
Target × Frequency × Phonetic Segment	−0.225	0.091	−2.477	0.013
Target × Frequency × Filter Type	−0.173	0.089	−1.949	0.051
Target × Phonetic Segment × Filter Type	−0.151	0.091	−1.652	0.098
Frequency × Phonetic Segment × Filter Type	−3.044	0.319	−9.556	<0.001
Target × Frequency × Phonetic Segment × Filter Type	0.025	0.176	0.142	0.887

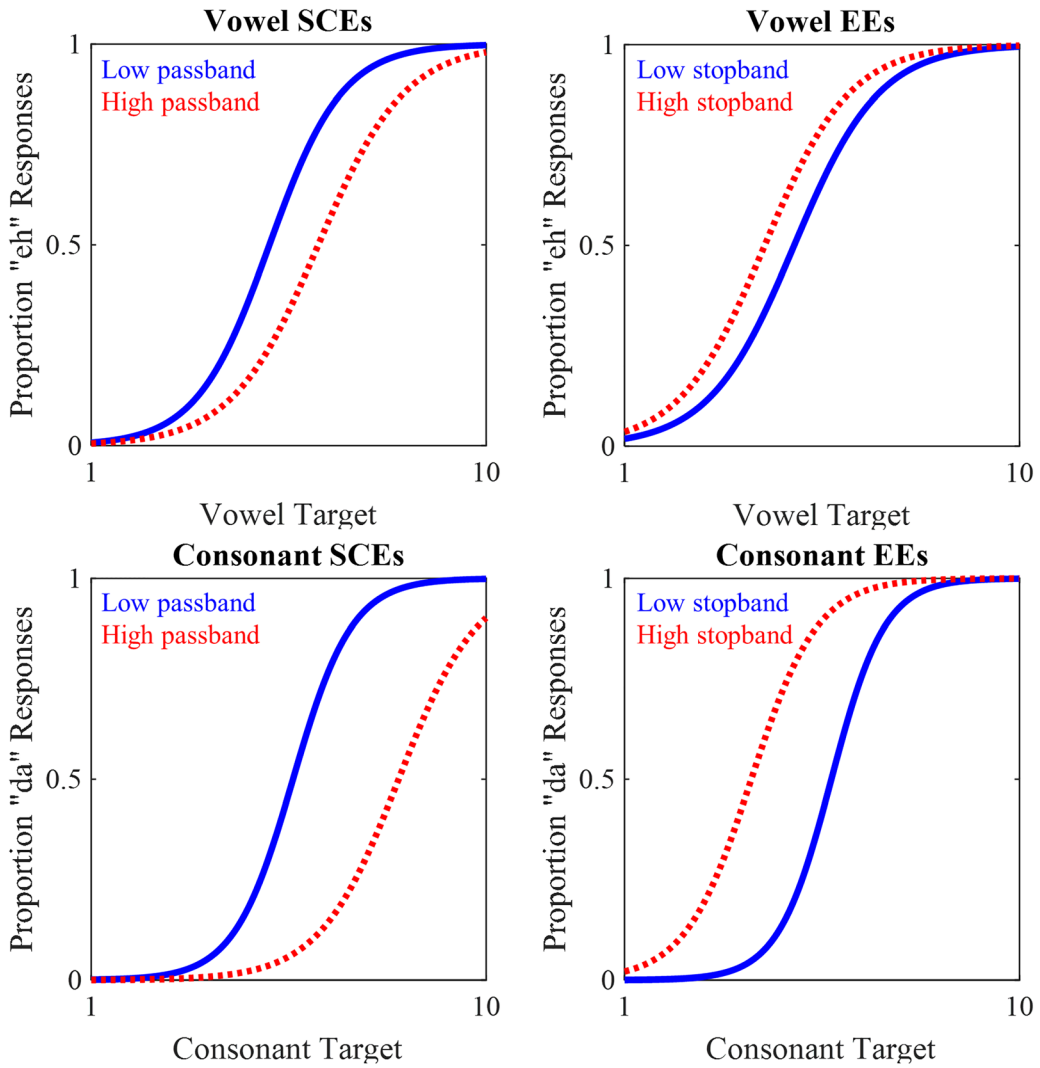

FIG. 2. (Color online) Results from Experiment 1. Mixed-effects models were fit to responses from each block individually for visualization. Blue solid lines depict responses following lower-frequency-modified context sentences. These functions are displaced to left when SCEs occurred (making higher-frequency "eh" / "da" responses more likely) and are displaced to the right when EEs occurred (making lower-frequency "ih" / "ga" responses more likely). Red dotted lines depict responses following higher-frequency-modified context sentences.

across Frequency). The Phonetic Segment factor by itself influenced listeners' responses (negative coefficient indicates listeners gave the higher-frequency response less often in the consonant task than in the vowel task), which might speak to the fact that the z-statistic was half as large in the Frequency by Phonetic Segment by Filter Type interaction ($z = -9.56$) compared to the Frequency by Filter Type interaction ($z = -19.75$).

Context effects magnitudes were calculated from these models following established procedures (Stilp *et al.*, 2015; Stilp and Assgari, 2017, 2018, 2019). First, a mixed-effects model was fit to responses in each of the four blocks individually. The main effects of Phonetic Segment and Filter Type and their related interactions dropped out (as each block tested one level for each factor), so each model tested the effects of Target, Frequency, the Target by Frequency interaction, random slopes for Target and Frequency, and a random intercept for listener. The 50% points were identified on the logistic regression fits to the aggregated responses following higher-frequency-filtered context and responses following lower-frequency-filtered contexts. These 50% points were then converted into the stimulus step number that listeners would label as the higher-frequency response option ("eh" or "da") 50% of the time. The 50% point on the higher-frequency-filtered regression function was derived as –Intercept/Slope in the regression model, and the 50% point on the lower-frequency-filtered function was derived as –(Intercept + Frequency)/(Slope + Target by Frequency). The context effect magnitude was defined as the distance between these 50% points, measured in the number of stimulus steps (interpolated as needed). The net shifts for SCEs were 0.67 stimulus steps (vowel task) and 1.82 stimulus steps (consonant task); the net shifts for EEs were −1.11 stimulus steps (vowel task) and −2.41 stimulus steps (consonant task) (Fig. 2).

Listeners' responses systematically varied beyond the influence of spectral context effects. The significant effect of Target indicates that the log-odds of responding with the higher-frequency target phoneme (/ɛ/ or /da/) increased with each rightward step along the target continuum (toward

higher-F_1 frequencies and the /ɛ/ endpoint in the vowel task; toward higher-F_3-onset frequencies and /da/ in the consonant task). Significant interactions with Target indicate that the slopes of psychometric functions varied as a function of experimental condition. Slopes were shallower in lower-frequency conditions than higher-frequency conditions (negative Target by Frequency interaction) and shallower in vowel categorization task than in the consonant categorization task (negative Target by Phonetic Segment interaction; see also negative three-way interaction between Target, Frequency, and Phonetic Segment). The trend towards a significant interaction between Target and Filter Type suggests modestly shallower slopes in the bandpass (SCE) conditions than in the bandstop (EE) conditions. This relationship exhibited trends towards negative interactions upon addition of the Frequency (shallower slopes in lower-frequency condition) or Phonetic Segment factors (shallower slopes for vowel conditions).

2. Individual differences analysis

Individual listeners' context effects were calculated by fitting generalized linear models to their responses in each block. These models included the predictors of Target and Frequency, as fitting models by-subject and by-block (which tested specific combinations of Filter Type and Phonetic Segment) eliminated the need for additional predictors. Some model fits that included Target × Frequency interactions were unstable, calculating context effects exceeding 100 steps in magnitude for some listeners, so this interaction term was not included in analyses. Models were fit to each listeners' responses in each block, and context effects were calculated as described above.

Context effect magnitudes for each listener are shown in scatterplots in Fig. 3. Four correlations of interest were calculated: vowel effects (SCE and EE), consonant effects (SCE and EE), SCEs (vowels and consonants), and EEs (vowels and consonants). The magnitudes of vowel context effects were not significantly correlated with each other ($r = 0.15$, $p = 0.60$), but consonant context effects were significantly correlated ($r = -0.70$, $p < 0.005$). When analyzing relationships by context effect type, EEs were not significantly correlated across frequency regions ($r = 0.16$, $p = 0.57$) but SCEs were significantly correlated across frequency regions ($r = 0.60$, $p < 0.025$).

C. Discussion

Experiment 1 explored the relationship between SCEs and EEs. Historically, SCEs in auditory perception have been produced by sentence context stimuli (though not exclusively; see Sec. I) and EEs have been generated by nonspeech context stimuli (multitone complexes, notched noise). Previous investigations of EEs in speech categorization used highly controlled nonspeech contexts to produce these effects (Summerfield *et al.*, 1984; Summerfield *et al.*, 1987; Thibodeau, 1991; Wang *et al.*, 2012), as did efforts to study SCEs and EEs within the same subject group (Coady *et al.*, 2003; Holt, 2006). Here, context sentences that produced SCEs in previous studies also produced EEs that biased

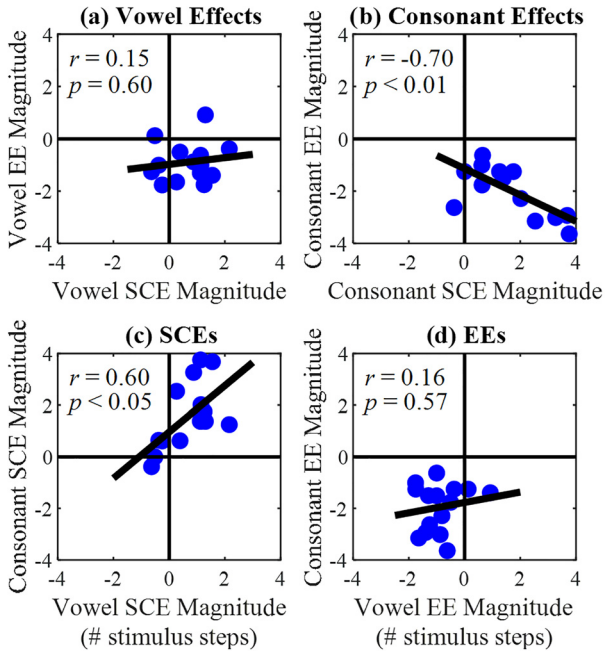

FIG. 3. (Color online) Scatterplots showing by-subject context effects in Experiment 1. All context effects are measured as the number of stimulus steps separating 50% points on that listener's (lower-frequency context and higher-frequency context) psychometric functions in that condition. Pearson's correlation coefficients and p-values are listed for each comparison. Solid lines depict linear regression fits to each data set.

phoneme categorization. These results demonstrate that EEs are not limited to carefully manipulated nonspeech context stimuli, but occur following highly acoustically variable speech contexts. Given these results and the generality of SCEs in auditory perception (Stilp *et al.*, 2015; Stilp and Assgari, 2019), it is quite possible that EEs also play a general and potentially widespread role in speech perception.

The individual differences analyses revealed some significant correlations among context effect magnitudes. SCE magnitudes were significantly positively correlated across frequency regions, as listeners who exhibited large SCEs in vowel categorization tended to exhibit large SCEs in consonant categorization as well [Fig. 3(c)]. Additionally, context effects were strongly correlated in the higher-frequency regions associated with consonant categorization: listeners who exhibited larger SCEs were likely to exhibit larger EEs for these stimuli as well [Fig. 3(b)]. The negative sign for this correlation reflects the fact that context effects were coded so that SCEs would have positive signs and EEs would have negative signs, reflecting categorization shifts in complementary directions. Other analyses of individual differences were not statistically significant: SCE and EE magnitudes were not correlated in the vowel categorization task [Fig. 3(a)], and EE magnitudes were not correlated across frequency regions [Fig. 3(d)]. Both of these null results include analyses of EEs in the vowel categorization task. This task introduced notches to the context sentence spectrum at relatively lower-frequency regions (100–400 and 550–850 Hz). These frequency regions, particularly the low-F_1 region, are lower than is typically studied in EE paradigms (which often test target frequencies in the vicinity of 1000 and 2000 Hz; e.g., Viemeister, 1980; Viemeister and

Bacon, 1982; Thibodeau, 1991; Byrne *et al.*, 2011, 2013; Kreft *et al.*, 2018). The ramifications of this discrepancy are discussed further following Experiment 2b. Nevertheless, the strong correlation between EE and SCE magnitudes at higher frequencies in the consonant categorization task offers concrete support for the suggested links between these effects (Holt and Lotto, 2002; Kluender *et al.*, 2003; Feng and Oxenham, 2018a).

Recently, Stilp *et al.* (2015), Stilp and Alexander (2016) and Stilp and Assgari (2017) discovered that SCEs are not all-or-none phenomena but that their magnitudes vary continuously. In those studies, filter gain was varied in 5-dB steps in order to add spectral peaks of varying magnitudes to the context sentence. SCE magnitudes increased linearly as a function of filter gain, revealing acute perceptual sensitivity to the size of the spectral difference across context and target stimuli. A few studies have considered the relative magnitudes of EE effects, showing that they vary as functions of precursor duration, interstimulus interval duration, and spectral notch width (Viemeister, 1980; Summerfield *et al.*, 1984; Viemeister *et* al., 2013; Feng and Oxenham, 2015; Kreft *et al.*, 2018). Yet, when considering energy at the target frequency in the context stimulus, most EE studies treat it in an all-or-none fashion: in simultaneous masking paradigms, energy is absent before the target and present in the target (or target-plus-masker complex); in forward masking paradigms, energy is absent in the adaptor and present in the masker and subsequent target. Summerfield *et al.* (1987) tested a more sensitive measure of how EEs shaped perception. Spectral notches in the precursor harmonic spectrum were located at frequencies corresponding to vowel formants, such that EEs would make the subsequent flat-spectrum harmonic spectrum sound like a vowel. As Summerfield *et al.* (1987) increased the notch depth in the context stimulus, vowel categorization became increasingly accurate, ostensibly due to progressively larger EEs. However, this is an indirect measure of EE magnitude, and effects saturated quickly (around 4–5 dB notch depth). Thus, how EE magnitudes vary as a function of the spectral difference between context and target is unclear. To further explore the relationship between SCEs and EEs, Experiment 2 introduced spectral notches of varying depths to the context sentence. If EE magnitudes are significantly correlated with notch depths, as SCE magnitudes are correlated with spectral peak magnitudes, this will deepen the relationship between these spectral context effects.

III. EXPERIMENT 2

A. Methods

1. Participants

Forty undergraduate students participated in Experiment 2 ($n = 20$ in Experiment 2a, $n = 20$ in Experiment 2b) in exchange for course credit. No one participated in multiple experiments. All self-reported being native English speakers with no known hearing impairments.

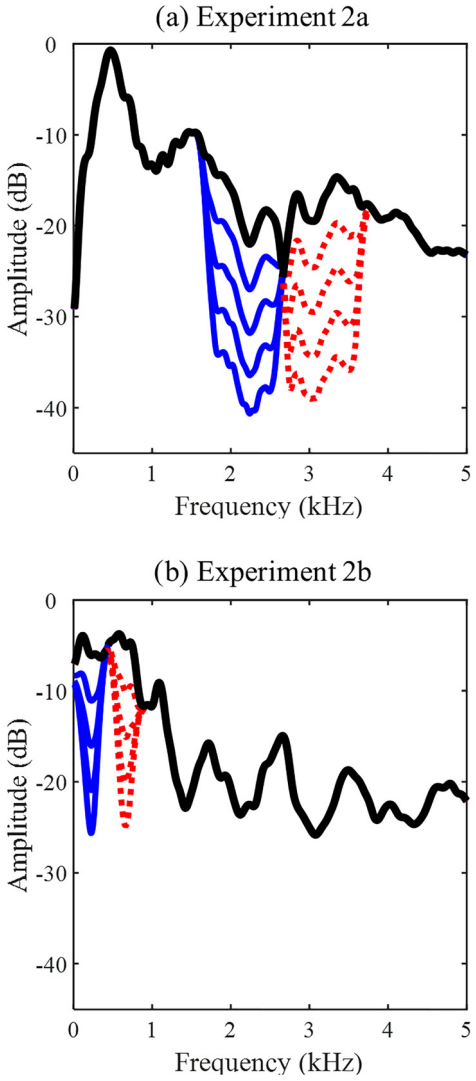

FIG. 4. (Color online) Long-term average spectra of sentence contexts from Experiment 2. Top: sentences with low-F_3 (1700–2700 Hz; solid blue lines) or high-F_3 (2700–3700 Hz; dotted red lines) spectral notches (−5, −10, −15, −20 dB). Bottom: sentences with low-F_1 (100–400 Hz; solid blue lines) or high-F_1 (550–850 Hz; dotted red lines) spectral notches (−5, −10, −15, −20 dB).

2. Stimuli

a. Contexts. Experiment 2 employed the same context sentences as in Experiment 1. In Experiment 2a, the context sentence "Correct execution of my instructions is crucial" was processed by bandstop filters with −5, −10, −15, or −20 dB gain in either the low-F_3 (1700–2700 Hz) or high-F_3 (2700–3700 Hz) region.[4] Filters were created using the fir2 command in MATLAB with 1000 coefficients and 5-Hz transitions regions. Importantly, these were the same filters as utilized in SCE experiments with variable-gain amplification of this context sentence (Stilp and Assgari, 2017), but here amplification was replaced by attenuation. In Experiment 2b, the context sentence "Please say what this vowel is" was processed by bandstop filters of varying depths. Bandstop attenuation was −5, −10, −15, or −20 dB in either the low-F_1 (100–400 Hz) or high-F_1 (550–850 Hz) region (Fig. 4). Filters were created using the fir2 command in MATLAB with 1000 coefficients and 5-Hz transition regions. As in

Experiment 2a, these were the same filters as utilized in SCE experiments with variable-gain amplification of this context sentence (Stilp *et al.*, 2015), but here with amplification replaced by attenuation. Following filtering, all contexts were low-pass filtered at a cutoff frequency of 5000 Hz.

b. Targets. Experiment 2a tested the same target consonants as described in Experiment 1, and Experiment 2b tested the same target vowels as described in Experiment 1. All context and target sounds were adjusted to produce equal rms amplitudes. Trial sequences were then created by concatenating one target to its corresponding context sentence with a 50-ms silent interstimulus interval. Finally, all stimuli were resampled at 44.1 kHz for presentation.

3. Procedure

Experiment 2 followed the same procedure as Experiment 1. As participants were only categorizing consonants (Experiment 2a) or vowels (Experiment 2b), each subgroup only completed the practice session that matched their main experiment. Each experiment consisted of four blocks (−5, −10, −15, or −20 dB filter gain), each comprised of 160 trials (2 frequency regions × 10 targets × 8 repetitions). Block order was again counterbalanced across listeners, and trials were presented in random orders within each block. The experiment was again self-paced. The entire experiment lasted approximately 40 min.

B. Results

The performance criterion of 80% correct on phoneme endpoints during the main experiment was again implemented, and all participants met this criterion. Results for Experiments 2a and 2b were analyzed using separate mixed-effects models with the same architectures (following Stilp and Assgari, 2018). The dependent variable was binary [lower-frequency responses ("ih," "ga") coded as 0 and higher-frequency responses ("eh," "da") coded as 1]. Fixed effects in the model included Target (coded as a continuous variable from 1 to 10 then mean-centered), Frequency (contrast coded with the higher-frequency region coded as −0.5 and the lower-frequency region coded as +0.5), Notch Depth (5, 10, 15, and 20 dB, mean-centered) and the interactions between fixed effects. Random slopes were included for each fixed main effect, and a random intercept of participant was also included.

Model results for Experiment 2a are reported in Table II and illustrated in Fig. 5.[3] As in Experiment 1, results directly pertaining to the primary research question will be addressed first. Consistent with Experiment 1, there was a significant negative effect of filter Frequency such that changing the spectral notch from the high-F_3 region (2700–3700 Hz, the condition coded −0.5) to the low-F_3 region (1700–2700 Hz, the condition coded +0.5) decreased the number of higher-frequency "da" responses. Thus, spectral notches in the context sentence need not be complete to produce EEs (as in Experiment 1). Critically, the Frequency by Notch Depth interaction was significant. This indicates that these variables shared a positive linear relationship such that EEs (the filter Frequency effect, which itself has a negative coefficient) increased in magnitudes at larger spectral notch depths (see Fig. 5).

TABLE II. Beta estimate (β), standard error (SE), Z statistic, and p value for fixed effects of the mixed-effects model for Experiment 2a. As described in the main text, Target and Notch Depth were each entered in the model as continuous factor centered around their respective means. Frequency was contrast-coded; the level associated with the −0.5 contrast is shown in parentheses.

Effect	β	SE	z	p
Intercept	0.230	0.206	1.117	0.264
Target	1.555	0.114	13.601	<0.001
Frequency (high F_3)	−1.676	0.198	−8.482	<0.001
Notch Depth	−0.037	0.013	−2.907	0.004
Target × Filter	−0.105	0.046	−2.276	0.023
Target × Notch Depth	−0.005	0.004	−1.105	0.269
Filter × Notch Depth	−0.113	0.020	−5.653	<0.001
Target × Filter × Notch Depth	−0.013	0.008	−1.558	0.119

Post hoc analyses were conducted to obtain a more sensitive test of the linear relationship between Frequency and Notch Depth. Following Stilp *et al.* (2015) and Stilp and Assgari (2017, 2018), Notch Depth was changed to a categorical factor. This selected one level of Notch Depth as the default level, then tested its model coefficient against 0 using a Wald z-test. All other model parameters matched those described in the above analysis. This process was repeated for all four levels of notch depth, and EE magnitudes were derived from the model each time using the same calculations described in Experiment 1 (separation of 50% points on logistic functions measured in number of stimulus steps). All EEs were significantly greater than 0 (all $z > -3.82$, $p < 0.001$), and EE magnitudes grew linearly with increasing notch depth ($r = -0.98$, $p < 0.025$; see regression fit in Fig. 5).

Additionally, the significant effect of Target reflects the increased log-odds of listeners responding "da" for each rightward step (towards higher F_3 onset frequencies and the /da/ endpoint) along the consonant continuum. The significant negative effect of Notch Depth indicates that listeners were less likely to respond "da" as spectral notch depth increased. Finally, the interaction between Target and Frequency was significant and negative, indicating shallower regression slopes for low-F_3-notched context sentences than high-F_3-notched context sentences (see Fig. 5).

Model results for Experiment 2b are reported in Table III and illustrated in Fig. 6. Results directly pertaining to the primary research question are again addressed first. As in Experiment 2a, there was a significant negative effect of filter Frequency such that changing the spectral notch from the high-F_1 region (550–850 Hz, the condition coded −0.5) to the low-F_1 region (100–400 Hz, the condition coded +0.5) decreased the number of higher-frequency "eh" responses. As in Experiment 2b, spectral notches in the context sentence need not be absolute in order to produce EEs. Critically, the Frequency by Notch Depth interaction was significant, suggesting a positive linear relationship to exist between EE magnitudes and spectral notch depths (see Fig. 6).

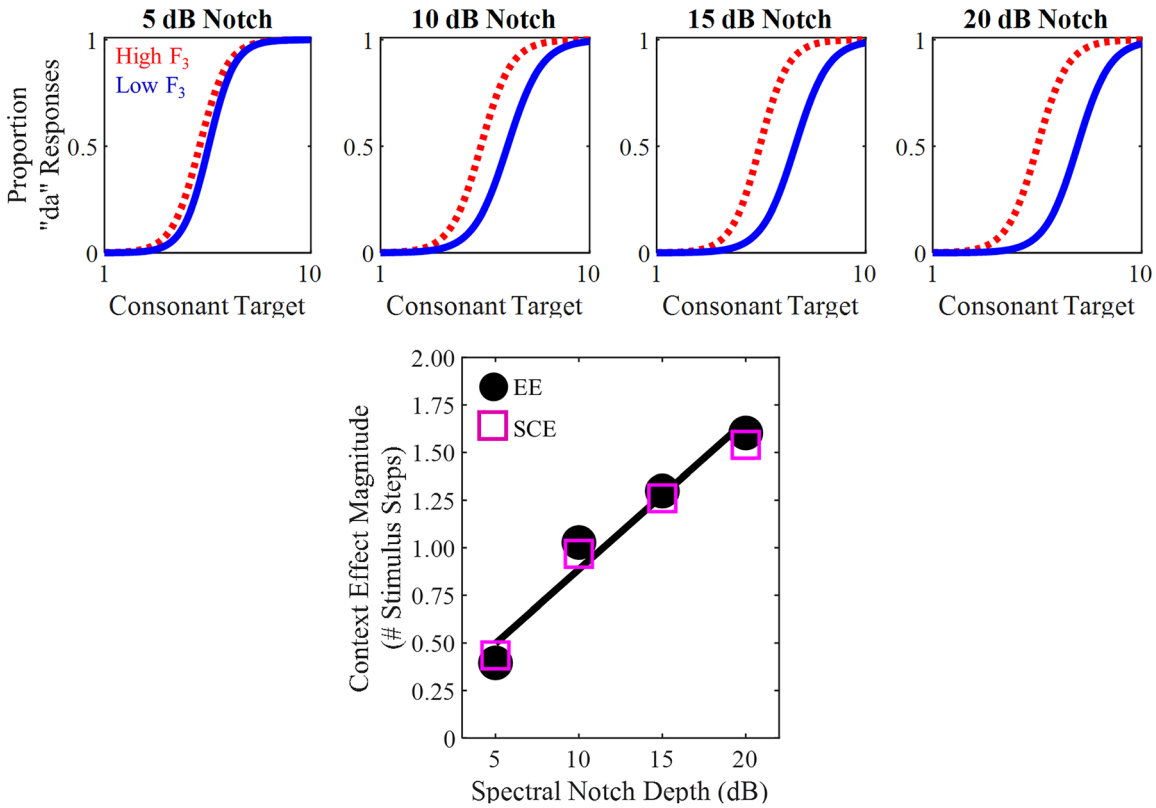

FIG. 5. (Color online) Results for Experiment 2a. The top row depicts mixed-effect models fit to each block of Experiment 2a at each level of spectral notch depth, with dotted red lines depicting responses to high-F_3-notch contexts and solid blue lines depicting responses to low-F_3-notch contexts. The bottom row shows the absolute magnitudes of EEs in Experiment 2a (black circles), the linear regression fit to these results (solid line), and SCE magnitudes from Stilp and Assgari (2017) using these stimuli and filters that amplified these spectral regions by comparable gains (5–20 dB).

Post hoc analyses were again conducted to obtain a more sensitive test of the linear relationship between filter Frequency and Notch Depth. Mixed-effect models were again fit to the data with Notch Depth as a categorical factor. This process was repeated for all four levels of notch depth, and EE magnitudes were derived from the model each time. All EEs were significantly greater than 0 (all $z > -3.69$, $p < 0.001$). However, unlike the results for Experiment 2a, here a strong linear relationship was not observed ($r = 0.82$, $p = 0.18$; see regression fit in Fig. 6). This is due in large part to EEs for smaller notch depths being of comparable magnitude (5 dB notch EE = −0.58 steps; 10 dB notch EE = −0.53 steps) and larger-notch-depth EEs being of comparable magnitude (15 dB notch EE = −0.98 steps; 20 dB notch EE = −0.92 steps), producing more of a step function than a linear function.

TABLE III. Beta estimate (β), standard error (SE), Z statistic, and p value for fixed effects of the mixed-effects model for Experiment 2b. As described in the main text, Target and Notch Depth were each entered in the model as continuous factor centered around their respective means. Frequency was contrast-coded; the level associated with the −0.5 contrast is shown in parentheses.

Effect	β	SE	Z	p
Intercept	0.103	0.104	0.988	0.323
Target	1.123	0.101	11.157	<0.001
Frequency (high F_1)	−0.848	0.128	−6.602	<0.001
Notch Depth	0.036	0.010	3.615	<0.001
Target × Filter	0.021	0.030	0.709	0.478
Target × Notch Depth	0.007	0.003	2.718	0.007
Filter × Notch Depth	−0.039	0.013	−3.040	0.002
Target × Filter × Notch Depth	−0.009	0.005	−1.694	0.090

Additionally, the significant effect of Target reflects the increased log-odds of listeners responding "eh" for each step to the right (towards higher F_1 frequencies) of the vowel continuum. The significant positive effect of Notch Depth indicates that listeners were more likely to respond "eh" as spectral notch depth increased. Finally, the interaction between Target and Notch Depth was significant, indicating steeper regression slopes as notch depth increased (see Fig. 6).

C. Discussion

In Experiment 2, EE magnitudes increased as a function of the depths of spectral notches in the context sentence. This parallels reports of SCE magnitudes increasing as a function of the magnitudes of spectral peaks added to the context sentences (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2017, 2019). In both cases, perceptual shifts increased to compensate for increasing spectral differences across context and target stimuli, revealing acute sensitivity to spectral changes across successive sounds. In conjunction with EE and SCE magnitudes in consonant categorization being correlated with each other in Experiment 1, the present results make a strong case for EEs and SCEs being more than just conceptually related to each other.

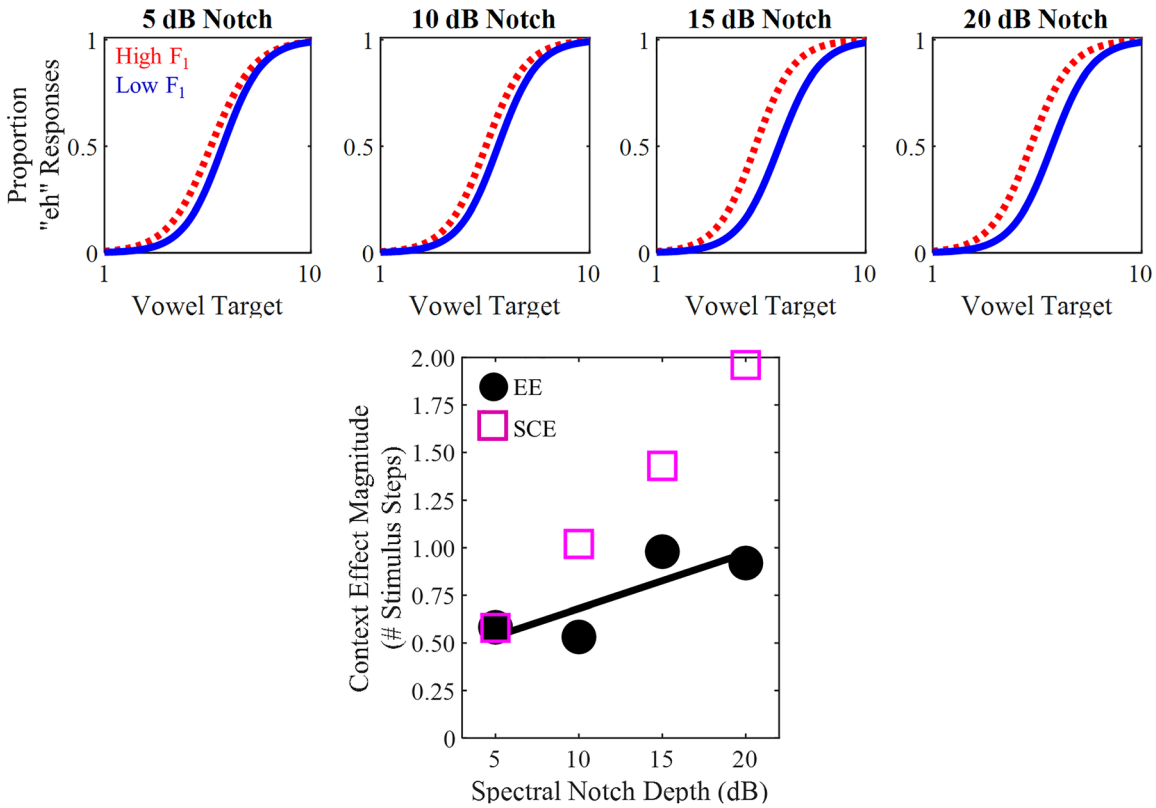

FIG. 6. (Color online) Results for Experiment 2b. The top row depicts mixed-effect models fit to each block of Experiment 2b at each level of spectral notch depth, with dotted red lines depicting responses to high-F_1-notch contexts and solid blue lines depicting responses to low-F_1-notch contexts. The bottom row shows the absolute magnitudes of EEs in Experiment 2b (black circles), the linear regression fit to these results (solid line), and SCE magnitudes from Stilp and Alexander (2016) using these stimuli and filters that amplified these spectral regions by comparable gains (5–20 dB).

Interactions in the mixed-effects models tested for linear relationships between fixed effects. Experiments 2a and 2b each exhibited significant interactions between Frequency and Notch Depth (Tables II and III), revealing that context effect magnitudes varied linearly as a function of spectral notch magnitude. This result comes from fitting one mixed-effects model to all data in each respective experiment, but this approach might suggest linearity to exist on a global scale that is not evident on a more local (block-by-block) scale. For example, Stilp and Assgari (2018) reported that SCE magnitudes varied linearly as a function of very small filter gains (+1 to +4 dB), but when analyzing results on a block-by-block level (without the model's inherent assumption that the relationship between variables is linear), context effects more closely resembled step functions than linear functions. Thus, *post hoc* analyses provided more sensitive tests of the linearity between spectral manipulations and subsequent categorization shifts.

In previous studies, spectral peaks added to context sentences and resulting SCEs were strongly linear at both global (omnibus mixed-effects models) and local levels (*post hoc* analyses matching those described above; Stilp *et al.*, 2015; Stilp and Assgari, 2017). This was also the case in Experiment 2a, as effects shared highly linear relationships in the omnibus model ($Z = -5.65$, $p < 0.0001$) and *post hoc* analyses ($r = -0.98$, $p < 0.025$). Furthermore, EE magnitudes at these spectral notch depths (from −5 to −20 dB in 5-dB-steps) closely correspond to SCE magnitudes in Stilp and Assgari (2017) at variable spectral peak magnitudes (from +5 to +20 dB in 5-dB-steps; see Fig. 5). Both studies employed the same context sentence and target stimuli, but used opposite spectral manipulations (spectral decrements versus increments) and produced categorization shifts in opposite directions (towards the decremented energy here; away from the increased energy in Stilp and Assgari, 2017). Nevertheless, EE magnitudes here and SCE magnitudes in Stilp and Assgari (2017) are highly correlated with each other ($r = -0.99$, $p < 0.01$), offering yet further support for a clear relationship between these context effects.

The results of Experiment 2b tell a different story. The omnibus model reported a significant Filter by Notch Depth interaction ($Z = -3.04$, $p < 0.01$), but *post hoc* analyses failed to support this linear relationship ($r = 0.82$, $p = 0.18$), instead looking more like a step function than a linear one (Fig. 6). This pattern of results also occurred in Stilp and Assgari (2018) for SCEs following very small spectral peaks added to the context sentence. Additionally, while the magnitudes of EEs in Experiment 2a closely resembled SCE magnitudes in a previous study, the same cannot be said for Experiment 2b. EEs were smaller and less linear than SCEs at comparable filter gains in Stilp and Alexander (2016) ($r = -0.80$, $p = 0.20$). The source of this discrepancy might be the investigation of EEs at lower (<1000 Hz) frequencies. As discussed following Experiment 1, past investigations of EEs have strongly tended to measure detection thresholds at higher frequencies (especially 1000 and 2000 Hz). The

neural mechanism thought to underlie EEs, adaptation of suppression/inhibition, would be available from frequencies both above and below the target frequency. In the present stimuli, a spectral notch in the low-F_1 region (100–400 Hz) does not receive much or potentially any suppression/inhibition from frequencies below it as there is little to no acoustic energy there. As such, perceptual shifts owing to adaptation of this suppressive/inhibitory influence would be irregular if not also quantitatively lesser compared to those occurring at higher frequencies (such as those tested in the consonant categorization task in Experiments 1 and 2a). Relatively lower frequencies exert greater masking on a higher frequency target than the other way around (Wegel and Lane, 1924), and adaptation of suppression/inhibition is more heavily influenced by frequencies below the target frequency than those above it (Carlyon, 1989; Nelson and Young, 2010). This asymmetry of suppression/inhibition might speak to why F_1 (vowel) EE magnitudes in Experiment 1 were not correlated with other context effects, but further research on this possibility is needed to be sure.

The frequency regions tested in Experiments 1 and 2 were inherited from previous studies of SCEs that used the same stimuli (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2017, 2018). As such, the bandwidths of spectral notches varied across Experiments 2a (1000 Hz) and 2b (300 Hz). EE magnitudes vary as a function of spectral notch bandwidth (Nelson and Young, 2010; Viemeister *et al.*, 2013; Kreft *et al.*, 2018). Here, while spectral notch bandwidths varied across experiments, the frequency regions also differed in their center frequencies. While notches in Experiment 2a had wider bandwidths in linear Hz, this difference dissipates somewhat when calculated in equivalent rectangular bandwidths (ERBs) (low-F_3 bandwidth = 3.88 ERBs, high-F_3 bandwidth = 2.73 ERBs, low-F_1 bandwidth = 6.03 ERBs, high-F_1 bandwidth = 3.03 ERBs). While EE magnitudes may vary for different notch bandwidths at these center frequencies, the central concern of Experiment 2 is how effects varied across different notch depths at fixed notch bandwidths.

IV. GENERAL DISCUSSION

The present investigation sought to clarify the relationship between two spectral context effects in auditory perception: SCEs and auditory EEs. In their respective literature, these effects have been studied in separate paradigms, with SCEs largely studied in speech categorization and EEs studied in detection of nonspeech stimuli (see Sec. I). Yet, both effects are demonstrations of enhanced perceptual sensitivity to spectral changes over time, leading investigators to suggest that these effects are related to one another (Holt and Lotto, 2002; Kluender *et al.*, 2003; Feng and Oxenham, 2018a). To date, this relationship has been only qualitative in nature, and attempts to link the two used unnatural context stimuli (Coady *et al.*, 2003; Holt, 2006).

The present experiments achieved the three goals outlined in the Introduction. First, SCEs and EEs were produced using the same sets of speech materials, as both biased categorization of vowels and consonants (Experiment 1). This provided a clear demonstration that acoustic contexts high in ecological validity and spectrotemporal variability can elicit EEs. Second, analyses of individual differences revealed that EE and SCE magnitudes in consonant categorization tasks were significantly correlated with each other. Third, EE magnitudes increased as a function of the depth of the spectral notches in the context sentence (Experiment 2), analogous to SCE magnitudes increasing as a function of the magnitude of the spectral peak added to the context sentence spectrum (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2017). Further, this increase was linear in nature in the consonant categorization task (Experiment 2a), replicating the linear relationships reported in previous studies with SCEs. Altogether, the present results support strong and clear links between these two context effects, well beyond the previously hypothesized qualitative relationship.

These results significantly broaden the potential role of EEs in speech perception. Previous investigations of EEs in speech perception used artificial context stimuli, such as notched harmonic spectra (Summerfield *et al.*, 1984; Coady *et al.*, 2003) or notched noise (Holt, 2006). Further, while Coady *et al.* (2003) and Holt (2006) reported EEs biasing perception of natural speech targets, investigations by Summerfield *et al.* (1984) and Summerfield *et al.* (1987) measured vowel recognition when EEs altered perception of a flat-harmonic-spectrum target. In all of these cases, context stimuli were severely constrained in terms of their spectrotemporal variability compared to the extraordinary variability of speech. The present reports used speech stimuli as both context and target, extending these context effects to stimuli commonly encountered in everyday listening. Also, EEs were observed following a wide range of spectral notch depths: from 5 to 20 dB (Experiment 2) to infinite notch depth (bandstop filtering; Experiment 1). Summerfield *et al.* (1987) reported progressive increases in vowel recognition (ostensibly produced by progressively larger EEs) as notch depths increased. The present results extend this relationship across both vowel and consonant categorization, as EE magnitudes increased as a function of notch depth in Experiment 2 (linearly in Experiment 2a, as a step function in Experiment 2b). Thus, spectral notches in context speech need not be complete in order to produce EEs (as in Experiment 1) but can be far more modest (as in Experiment 2). Speech contexts can naturally possess spectral compositions that influence sound categorization via SCEs (Stilp and Assgari, 2019); while it is an empirical question that the same be true for EEs, the present results make that a likely proposition. While SCEs have been proposed to have a prominent and widespread influence of speech perception (Stilp *et al.*, 2015; Stilp and Assgari, 2019), EEs might also be characterized by similar generality and pervasiveness.

Neural mechanisms related to adaptation are thought to underlie both of these spectral context effects. Specifically, SCEs are thought to be produced by simple neural adaptation and EEs are thought to be produced by neural adaptation of suppression/inhibition (see Sec. I). With regard to SCEs, results from Experiment 1 are consistent with this proposal. For example, the low-F_1 passband context would activate and subsequently adapt neurons at those frequencies.

Neurons responding to higher-F_1 frequencies would be less adapted or entirely unadapted by this context stimulus, making them relatively more responsive to the following target stimulus than low-F_1 neurons would be. This neural contrast would underlie the increase in high-F_1 ("eh") responses to the target vowel. With regard to EEs, results from both experiments are entirely consistent with its proposed mechanism. Suppression and/or inhibition of target frequencies were (at least partially) adapted by the end of the context sentence, making them perceptually more prominent upon introduction of the target stimulus, producing an increase in responses at those frequencies. This occurred when spectral energy in that target frequency region was entirely absent (Experiment 1) or only relatively reduced (Experiment 2). However, the present results shed at least some doubt on the possibility that SCEs are also produced by adaptation of suppression/inhibition. The use of spectrally narrowband contexts in Experiment 1 would produce minimal suppression/inhibition from neighboring frequencies onto the target frequencies. Previous studies have reported SCEs being produced by spectral contexts of extremely limited bandwidth, even as narrow as a single pure tone (Lotto and Kluender, 1998). It is difficult to reconcile those results with suppression/inhibition from neighboring frequencies (where no energy was present) when adaptation produced by energy in the context stimulus provides a more parsimonious explanation. The present results do not conclusively demonstrate the neural mechanisms underlying these effects, but do offer some suggestion of SCEs and EEs being produced by related mechanisms. Another possibility exists wherein these effects emanate from the same underlying processes.[5] Adapting a given channel may produce one effect (e.g., a decrease in neural activity, as in SCEs) but that channel inhibiting adjacent channels at later levels of processing may produce a different effect (e.g., an increase in neural activity, as in EEs). Further behavioral and physiological data are needed to distinguish these possibilities.

Considerable debate has surrounded where these context effects occur in the auditory system. Central to this debate is the distinction whether effects are more peripheral (cochlear) or central (post-cochlear) in nature. For SCEs, this question was first addressed by Watkins (1991), who presented context and target stimuli monotically, diotically, and dichotically (context presented in one ear followed by the target presented in the opposite ear). SCEs occurred for dichotic stimulus presentations, confirming that SCEs are not purely peripheral but do receive contributions from central processing. However, SCE magnitudes were substantially diminished when stimuli were presented dichotically compared to monotic or diotic presentations. This offers some suggestion that SCEs are not produced exclusively by central processing but occur both peripherally and centrally (see Holt and Lotto, 2002 and Feng and Oxenham, 2018b for similar results). For EEs, several studies reported failures to produce these effects under dichotic stimulus presentation (e.g., Viemeister, 1980; Summerfield *et al.*, 1987; Carlyon, 1989), leading to suggestions that EEs originate in the auditory periphery. More recently, multiple reports offered positive evidence of EEs being produced by contralateral (dichotic) stimulus presentation, which supports contributions from central processing to these effects (Erviti *et al.*, 2011; Kidd *et al.*, 2011; Carcagno *et al.*, 2012; Byrne *et al.*, 2013). Physiological data explaining SCEs are not yet available, but such measures have been reported for EEs. Auditory nerve fibers in anesthetized animals do not display evidence of auditory enhancement (Palmer *et al.*, 1995), but neural responses consistent with EEs have been reported in the inferior colliculus of awake animals (Nelson and Young, 2010). Given the presence of adaptation-related mechanisms throughout the auditory system, these effects might not be restricted to a single locus but emerge and/or repeat at successive levels of the auditory system (Nelson and Young, 2010; Carcagno *et al.*, 2012; Feng and Oxenham, 2015).

Recent reports have demonstrated that these spectral context effects are not limited to healthy hearing. SCEs biased vowel categorization for listeners with sensorineural hearing loss (Stilp and Alexander, 2016) and cochlear implant users (Feng and Oxenham, 2018b; see also Stilp, 2017). In all three reports, the magnitudes of SCEs were larger than those observed for normal-hearing listeners. This is potentially problematic because larger-than-normal SCEs can impair perceptual accuracy (Stilp, 2017). Similarly, EEs have been observed for listeners with sensorineural hearing loss (Thibodeau, 1991; Kreft *et al.*, 2018) as well as cochlear implant users (Goupell and Mostardi, 2012; Wang *et al.*, 2012; Kreft and Oxenham, 2017; Feng and Oxenham, 2018a). In these cases, EE magnitudes were smaller than those reported for normal-hearing listeners. Additionally, while normal-hearing listeners experience EEs in forward masking and simultaneous masking paradigms, cochlear implant users only experienced (diminished) EEs in the latter case (Kreft and Oxenham, 2017). While the present results support a clear relationship between SCEs and EEs, they might require different strategies in assistive listening devices if the goal is to approximate the respective magnitudes of normal-hearing listeners' context effects (i.e., to decrease the magnitudes of SCEs and increase the magnitudes of EEs).

While supporting a clear connection between SCEs and EEs in speech perception, the present studies are clearly not the only ways to test this relationship. There exists a large parameter space through which these effects have been studied and their similarity can be further assessed. For example, both effects persist across sizable interstimulus intervals. Viemeister (1980) reported small EEs for detection of 80-ms 1-kHz targets following 2400-ms multitone complex adaptors, 100-ms maskers, and 6400-ms interstimulus intervals (ISIs). Broadbent *et al.* (1956) reported that half of their listeners still exhibited response shifts (consistent with SCEs) when the context sentence and target vowel were separated by a 10-s ISI. But, strong conclusions about how these results inform the EE-SCE relationship must be tempered by the sizable acoustic differences across stimuli. Along with Coady *et al.* (2003) and Holt (2006), the present experiments demonstrated that variants of the same stimuli can be used to study both EEs and SCEs. This offers a platform for future research where controlled manipulation of various acoustic and experimental parameters (interstimulus intervals,

relative amplitudes of context and targets, durations, context spectral feature prominence as in Experiment 2, etc.) will further define the relationship between these effects of surrounding spectral context on perception.

ACKNOWLEDGMENTS

The author thanks Ginny Richards, Andrew Oxenham, and two reviewers for their insightful feedback and suggestions. The author also thanks Scott Barrett, Ella Beilman, Samantha Cardenas, Rebecca Davis, Emily Dickey, Jonathan Frazier, Joshua Lanning, Caroline Smith, Pratistha Thapa, and Sara Wardrip for their assistance with data collection.

[1]Previous studies with these consonant stimuli used a different coding scheme, where the continuum progressed from step 1 (high-F_3-onset endpoint /da/) to 10 (low-F_3-onset endpoint /ga/), the Filter effect had a default level of high F_3 amplification, and the lower-frequency response option /ga/ was coded as 1. This produced SCEs with a negative effect of Filter, as changing the filtering condition from high F_3 to low F_3 resulted in a significant decrease in lower-frequency /ga/ responses (and more higher-frequency /da/ responses). This shift was in the opposite direction as those observed for vowel sounds, where the continuum arranged from low-F_1 endpoint /ɪ/ to high-F_1 endpoint /ɛ/, the default level of Filter was high-F_1 amplification, and the higher-frequency response option /ɛ/ was coded as 1 (Stilp and Assgari, 2018). In that coding scheme, Filter exerted a significant positive effect on responses, as changing the filtering from high F_1 to low F_1 significantly increased high-F_1 responses. Here, the consonant continuum has been arranged from low-F_3-onset endpoint to high-F_3-onset endpoint, the default level of Filter is again high-F_3 amplification, and the higher-frequency response option is coded with 1. Matching coding schemes across vowel and consonant stimuli facilitates comparisons across context effect types (SCEs producing positive effects of Filter, EEs exhibiting negative effects of Filter).

[2]Setting bandpass filtering (which is predicted to produce SCEs) as the default level of the Filter Type variable reflects the fact that these stimuli have produced SCEs in past studies (Stilp *et al.*, 2015; Stilp and Alexander, 2016; Stilp and Assgari, 2017, 2018, 2019). The present study is the first test of whether the same stimuli might also produce EEs.

[3]Data and analysis scripts for all experiments are available at https://osf.io/m8647.

[4]Following nonlinear processes such as cochlear compression, one would not predict that these notch depths are veridically maintained in internal stimulus representations in the auditory system. Such veridical recovery is not necessary here, as the objective of the experiment is to test similar ranges of spectral modifications to context sentences as tested in previous studies of SCEs (Stilp and Alexander, 2016; Stilp and Assgari, 2017).

[5]We wish to thank an anonymous reviewer for raising this point.

Bates, D. M., Maechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and S4. R package (version 1.1-7) [computer program]," http://cran.r-project.org/package=lme4 (Last viewed 7/25/19).

Boersma, P., and Weenink, D. (**2017**). "Praat: Doing phonetics by computer (version 6.0.36) [computer program]," http://www.praat.org (Last viewed 7/25/19).

Broadbent, D. E., Ladefoged, P., and Lawrence, W. (**1956**). "Vowel sounds and perceptual constancy," Nature **178**(4537), 815–816.

Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (**2011**). "The enhancement effect: Evidence for adaptation of inhibition using a binaural centering task," J. Acoust. Soc. Am. **129**(4), 2088–2094.

Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (**2013**). "The salience of enhanced components within inharmonic complexes," J. Acoust. Soc. Am. **134**(4), 2631–2634.

Carcagno, S., Semal, C., and Demany, L. (**2012**). "Auditory enhancement of increments in spectral amplitude stems from more than one source," J. Assoc. Res. Otorthinolaryngol. **13**(5), 693–702.

Carlyon, R. P. (**1989**). "Changes in the masked thresholds of brief tones produced by prior bursts of noise," Hear. Res. **41**(2–3), 223–235.

Coady, J. A., Kluender, K. R., and Rhode, W. S. (**2003**). "Effects of contrast between onsets of speech and other complex spectra," J. Acoust. Soc. Am. **114**(4), 2225–2235.

Delgutte, B. (**1996**). "Auditory neural processing of speech," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell Publishing Ltd., Oxford, UK), pp. 507–538.

Delgutte, B., Hammond, B. M., Kalluri, S., Litvak, L. M., and Cariani, P. A. (**1996**). "Neural encoding of temporal envelope and temporal interactions in speech," in *Proceedings of Auditory Basis of Speech Perception*, July 15–19, Keele, UK.

Erviti, M., Semal, C., and Demany, L. (**2011**). "Enhancing a tone by shifting its frequency or intensity," J. Acoust. Soc. Am. **129**(6), 3837–3845.

Feng, L., and Oxenham, A. J. (**2015**). "New perspectives on the measurement and time course of auditory enhancement," J. Exp. Psychol. Human Percept. Perform. **41**(6), 1696–1708.

Feng, L., and Oxenham, A. J. (**2018a**). "Auditory enhancement and the role of spectral resolution in normal-hearing listeners and cochlear-implant users," J. Acoust. Soc. Am. **144**(2), 552–566.

Feng, L., and Oxenham, A. J. (**2018b**). "Effects of spectral resolution on spectral contrast effects in cochlear-implant users," J. Acoust. Soc. Am. **143**(6), EL468–EL473.

Feng, L., and Oxenham, A. J. (**2018c**). "Spectral contrast effects produced by competing speech contexts," J. Exp. Psychol. Human Percept. Perform. **44**(9), 1447–1457.

Frazier, J., Assgari, A. A., and Stilp, C. E. (**2019**). "Musical instrument categorization is highly sensitive to spectral properties of earlier sounds," Atten. Percept. Psychophys. **81**(4), 1119–1126.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," NIST Order No. PB91-505065, National Institute of Standards and Technology, Gaithersburg, MD.

Goupell, M. J., and Mostardi, M. J. (**2012**). "Evidence of the enhancement effect in electrical stimulation via electrode matching (L)," J. Acoust. Soc. Am. **131**(2), 1007–1010.

Holt, L. L. (**2005**). "Temporally nonadjacent nonlinguistic sounds affect speech categorization," Psychol. Sci. **16**(4), 305–312.

Holt, L. L. (**2006**). "The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization," J. Acoust. Soc. Am. **120**(5), 2801–2817.

Holt, L. L., and Lotto, A. J. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**(1–2), 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**(2), 710–722.

Kidd, G., Richards, V. M., Streeter, T., Mason, C. R., and Huang, R. (**2011**). "Contextual effects in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **130**(6), 3926–3938.

Kidd, G., and Wright, B. A. (**1994**). "Improving the detectability of a brief tone in noise using forward and backward masker fringes: Monotic and dichotic presentations," J. Acoust. Soc. Am. **95**(2), 962–967.

Kingston, J., Kawahara, S., Chambless, D., Key, M., Mash, D., and Watsky, S. (**2014**). "Context effects as auditory contrast," Atten. Percept. Psychophys. **76**, 1437–1464.

Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**(1), 59–69.

Kreft, H. A., and Oxenham, A. J. (**2017**). "Auditory enhancement in cochlear-implant users under simultaneous and forward masking," J. Assoc. Res. Otolaryngol. **18**(3), 483–493.

Kreft, H. A., Wojtczak, M., and Oxenham, A. J. (**2018**). "Auditory enhancement under simultaneous masking in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **143**(2), 901–910.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**(1), 98–104.

Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**(4), 602–619.

Nelson, P. C., and Young, E. D. (**2010**). "Neural correlates of context-dependent perceptual enhancement in the inferior colliculus," J. Neurosci. **30**(19), 6577–6587.

Palmer, A. R., Summerfield, Q., and Fantini, D. A. (**1995**). "Responses of auditory-nerve fibers to stimuli producing psychophysical enhancement," J. Acoust. Soc. Am. **97**(3), 1786–1799.

R Development Core Team. (**2016**). "R: A language and environment for statistical computing," Vienna, Austria: R Foundation for Statistical Computing, http://www.r-project.org/ (Last viewed 7/25/19).
Schouten, J. (**1940**). "The residue and the mechanism of hearing," Proc. K. Ned. Akad. Wet. **43**, 991–999.
Sjerps, M. J., Zhang, C., and Peng, G. (**2018**). "Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context," J. Exp. Psychol. Human Percept. Perform. **44**(6), 914–924.
Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., and Cook, S. (**2012**). "Development and validation of the AzBio sentence lists," Ear Hear. **33**(1), 112–117.
Stephens, J. D. W., and Holt, L. L. (**2011**). "A standard set of American-English voiced stop-consonant stimuli from morphed natural speech," Speech Commun. **53**(6), 877–888.
Stilp, C. E. (**2017**). "Acoustic context alters vowel categorization in perception of noise-vocoded speech," J. Assoc. Res. Otolaryngol. **18**(3), 465–481.
Stilp, C. E., and Alexander, J. M. (**2016**). "Spectral contrast effects in vowel categorization by listeners with sensorineural hearing loss," Proc. Mtgs. Acoust. **26**, 060003.
Stilp, C. E., Alexander, J. M., Kiefte, M., and Kluender, K. R. (**2010**). "Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets," Atten. Percept. Psychophys. **72**(2), 470–480.
Stilp, C. E., Anderson, P. W., and Winn, M. B. (**2015**). "Predicting contrast effects following reliable spectral properties in speech perception," J. Acoust. Soc. Am. **137**(6), 3466–3476.
Stilp, C. E., and Assgari, A. A. (**2017**). "Consonant categorization exhibits a graded influence of surrounding spectral context," J. Acoust. Soc. Am. **141**(2), EL153–EL158.
Stilp, C. E., and Assgari, A. A. (**2018**). "Perceptual sensitivity to spectral properties of earlier sounds during speech categorization," Atten. Percept. Psychophys. **80**(5), 1300–1310.
Stilp, C. E., and Assgari, A. A. (**2019**). "Natural signal statistics shift speech sound categorization," Atten. Percept. Psychophys. **81**(6), 2037–2052.
Summerfield, Q., Haggard, M., Foster, J., and Gray, S. (**1984**). "Perceiving vowels from uniform spectra—Phonetic exploration of an auditory aftereffect," Percept. Psychophys. **35**(3), 203–213.
Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). "Auditory enhancement of changes in spectral amplitude," J. Acoust. Soc. Am. **81**(3), 700–708.
Thibodeau, L. M. (**1991**). "Performance of hearing-impaired persons on auditory enhancement tasks," J. Acoust. Soc. Am. **89**(6), 2843–2850.
Viemeister, N. F. (**1980**). "Adaptation of masking," in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by G. V. D. Brink and F. A. Bilsen (University Press, Delft, the Netherlands), pp. 190–198.
Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**(6), 1502–1507.
Viemeister, N. F., Byrne, A. J., and Stellmack, M. A. (**2013**). "Spectral and level effects in auditory signal enhancement," in *Basic Aspects of Hearing* (Springer, New York), pp. 167–174.
Wang, N. Y., Kreft, H., and Oxenham, A. J. (**2012**). "Vowel enhancement effects in cochlear-implant users," J. Acoust. Soc. Am. **131**(6), EL421–EL426.
Wang, N., and Oxenham, A. J. (**2016**). "Effects of auditory enhancement on the loudness of masker and target components," Hear. Res. **333**, 150–156.
Watkins, A. J. (**1991**). "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **90**(6), 2942–2955.
Wegel, R. L. F., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**(2), 266–285.
Winn, M. B., and Litovsky, R. Y. (**2015**). "Using speech sounds to test functional spectral resolution in listeners with cochlear implants," J. Acoust. Soc. Am. **137**(3), 1430–1442.
Winn, M. B., Rhone, A. E., Chatterjee, M., and Idsardi, W. J. (**2013**). "The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants," Front. Psychol. **4**, 824.