Hearing Research 392 (2020) 107983

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Evaluating peripheral versus central contributions to spectral context effects in speech perception

Christian E. Stilp

Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, 40292, USA

ARTICLE INFO

Article history:
Received 14 February 2020
Received in revised form
7 April 2020
Accepted 28 April 2020
Available online 15 May 2020

Keywords:
Enhancement effect
Contrast effect
Context effect
Speech perception
Auditory perception
Perceptual constancy

ABSTRACT

Perception of a sound is influenced by spectral properties of surrounding sounds. When frequencies are absent in a preceding acoustic context before being introduced in a subsequent target sound, detection of those frequencies is facilitated via an auditory enhancement effect (EE). When spectral composition differs across a preceding context and subsequent target sound, those differences are perceptually magnified and perception shifts via a spectral contrast effect (SCE). Each effect is thought to receive contributions from peripheral and central neural processing, but the relative contributions are unclear. The present experiments manipulated ear of presentation to elucidate the degrees to which peripheral and central processes contributed to each effect in speech perception. In Experiment 1, EE and SCE magnitudes in consonant categorization were substantially diminished through contralateral presentation of contexts and targets compared to ipsilateral or bilateral presentations. In Experiment 2, spectrally complementary contexts were presented dichotically followed by the target in only one ear. This arrangement was predicted to produce context effects peripherally and cancel them centrally, but the competing contralateral context minimally decreased effect magnitudes. Results confirm peripheral and central contributions to EEs and SCEs in speech perception, but both effects appear to be primarily due to peripheral processing.

1. Introduction

Perception of a sound is highly sensitive to the spectral properties of surrounding sounds. This is typified by two classic effects of surrounding spectral context in auditory perception. First, the salience and/or detectability of frequencies is shaped through auditory enhancement effects (EEs) (Schouten, 1940; Viemeister, 1980; Viemeister and Bacon, 1982; Byrne et al., 2011; Erviti et al., 2011; Carcagno et al., 2012; Feng and Oxenham, 2018a). For example, when EEs are measured in a simultaneous masking paradigm, the target frequency is often embedded in a multitone complex. Relative to presenting this complex in isolation, target frequency detection thresholds are improved when this masker-plus-target complex follows an adaptor stimulus with the same spectrum minus the target frequency. When EEs are measured in a forward masking paradigm, detection thresholds for a tone are better when it follows only a masker stimulus compared to when an adaptor stimulus with the same spectrum as the masker minus energy at the target frequency precedes (and enhances) the masker. A wide variety of experimental tasks have produced EEs, including binaural centering tasks (Byrne et al., 2011), pitch salience judgments (Byrne et al., 2013), pitch movement judgments (Erviti et al., 2011; Carcagno et al., 2012; Feng and Oxenham, 2015), and both vowel and consonant categorization (Summerfield et al., 1984, 1987; Coady et al., 2003; Holt, 2006; Stilp, 2019).

Second, perception of complex sounds is also shaped by spectral contrast effects (SCEs) (Ladefoged and Broadbent, 1957; Watkins, 1991; Stilp et al., 2010; Stilp et al., 2015; Sjerps and Reinisch, 2015; Feng and Oxenham, 2018b; Sjerps et al., 2018). In a seminal paper, Ladefoged and Broadbent (1957) demonstrated that the spectrum of a preceding sentence context influenced categorization of the subsequent vowel targets. When the first formant frequencies (F_1) of the context sentence were raised to higher frequencies, listeners perceived the subsequent target vowel as /ɪ/ (as in "bit", low F_1) more often. When the F_1 frequencies of the context sentence were lowered, listeners perceived the target as /ɛ/ (as in "bet", high F_1) more often. Considerable research has demonstrated the pervasiveness of SCEs in perception of many different vowels and consonants (for review see Stilp, 2020). These

E-mail address: christian.stilp@louisville.edu.

https://doi.org/10.1016/j.heares.2020.107983
0378-5955/

effects extend to categorization of nonspeech sounds as well (Stilp et al., 2010; Kingston et al., 2014; Lanning and Stilp, 2020).

EEs and SCEs are both thought to be produced by processes related to neural adaptation. Simple neural adaptation has been offered as a plausible mechanism underlying SCEs (e.g., Delgutte, 1996; Delgutte et al., 1996; Holt et al., 2000; Holt and Lotto, 2002; Stilp, 2019). Frequencies in the context would adapt neurons coding those frequencies, making them less responsive when the target sound is introduced. Neurons coding neighboring frequencies would be unadapted/less adapted and thus relatively more responsive to the frequencies in the target. This neural contrast would underlie a perceived shift in spectral content of the target, producing an SCE. Adaptation of inhibition has been proposed as the mechanism underlying EEs (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). Neurons responding to frequencies in the context also inhibit neurons encoding neighboring frequencies, but this inhibition adapts over time. This results in neural responses to inhibited frequencies being more pronounced than they were initially, such as when frequencies absent in the context are introduced in the target. Here 'adaptation of inhibition' is used as the broad label for the hypothesized underlying mechanism that produces EEs, for which adaptation of suppression in the cochlea is taken as a specific instance.

Determining the neural locus/loci of these effects has been challenging, as adaptation and adaptation of inhibition both occur at several sites in the ascending auditory system. Direct neural evidence has been reported for SCEs occurring in human primary auditory cortex (Sjerps et al., 2019), and for EEs occurring in the inferior colliculus of marmosets (Nelson and Young, 2010) and both subcortical and cortical responses of humans measured via EEG (Feng et al., 2018). However, these results do not distinguish whether effects first occurred at these nuclei or were instead inherited from earlier (lower) neural processing. Palmer, Summerfield, and Fantini (1995) did not find evidence of EEs in the auditory nerve fibers of guinea pigs, but their use of anesthesia complicates interpretation of this null result. Scutt and Palmer (1998) reported enhancement of tones in the cochlear nucleus, but effects were limited to tone onsets. Several investigators have raised the possibility that these effects might not be restricted to a single neural locus but instead repeat at successive levels of the auditory system (Nelson and Young, 2010; Carcagno et al., 2012; Feng and Oxenham, 2015; Feng et al., 2018).

While behavioral experiments cannot conclusively establish neural mechanisms, investigations that varied the ear(s) of presentation have provided valuable insights to the geneses of SCEs and EEs. Presenting the context and subsequent target stimuli monotically limits the unique contribution of central processing[1] to the effect since it has already manifest peripherally. Conversely, presenting these stimuli contralaterally (i.e., the context presented to one ear followed by the target presented to the opposite ear) restricts the contributions from peripheral processing, as neither ear receives both context and target stimuli. For SCEs, ear of presentation was first manipulated by Watkins (1991), who presented contexts filtered to emphasize spectral properties of the following target vowels (/ɪ/ as in "bit" and /ɛ/ as in "bet"). Ipsilateral presentation of the sentence contexts and target vowels produced SCEs; contralateral presentation significantly reduced SCE magnitudes but the effect still occurred. This led Watkins to suggest that SCEs were mediated by central mechanisms. When repeating the experiments using signal-correlated noise carriers that shared the amplitude envelope and spectral envelope of speech carriers, ipsilateral presentation still produced an SCE but contralateral presentation extinguished the effect altogether. This pattern of results was instead consistent with a peripheral mechanism. From these and other results, Watkins (1991) concluded that peripheral mechanisms alone cannot explain SCEs, but that effects were primarily produced by central mechanisms. Feng and Oxenham (2018b) used similar methods to Watkins (1991) and also observed smaller SCEs for contralateral presentation of context and target stimuli relative to ipsilateral presentation, but concluded that peripheral processing appeared to dominate the effect. Holt and Lotto (2002) investigated the effect of mode of presentation on SCEs using short-term contexts (/i/ and /u/ contexts biasing perception of /ba/ and /da/ targets; /al/ and /ar/ contexts biasing perception of /da/ and /ga/ targets). In both cases, diotic and dichotic stimulus presentation produced SCEs. For syllable contexts, SCEs in /da/-/ga/ categorization were significantly larger in diotic presentation than dichotic presentation. For vowel contexts, however, SCEs in /ba/-/da/ categorization were marginally larger in dichotic presentation. Holt and Lotto concluded that peripheral processing was facilitative for producing SCEs but not solely responsible, instead attributing these effects to be primarily central in nature. Results in the /da/-/ga/ categorization task were replicated by Lotto et al. (2003) when the context syllables /al/ and /ar/ were replaced by tone analogues that mimicked the time-varying frequencies of F_3 in the syllables. The authors noted the possibility that peripheral mechanisms play at least some role in producing these SCEs, but again classified these effects as primarily central.

Early reports of EEs in auditory perception were produced using ipsilateral presentation of context and target stimuli; failures to observe EEs in contralateral presentations led to suggestions that EEs originated in the auditory periphery (e.g., Viemeister, 1980; Summerfield et al., 1987; Carlyon, 1989). More recent reports offered positive evidence of EEs being produced by contralateral presentation of context and target stimuli. Erviti, Semal, and Demany (2011) presented trials with four inharmonic tone complexes (alternating between precursor and target stimuli) followed by a probe tone. Precursor chords differed from target chords by having a slightly lower intensity for one component (producing an intensity EE in perception of the target chord) or a slightly different frequency for one tone (producing a frequency EE in perception of the target chord). Listeners reported whether the probe tone was present or absent in the target chords. Intensity EEs occurred but with smaller magnitudes in contralateral presentation relative to ipsilateral presentation; frequency EEs were comparable across both modes of presentation. Dichotic intensity EEs were also reported by Carcagno et al. (2012), who used the same task but only one presentation apiece of the precursor and target per trial. The target frequency in inharmonic complexes was enhanced by ipsilateral and contralateral EEs, both of which maintained across varying interstimulus intervals and whether the probe tone followed or preceded the precursor and target stimuli. Kidd et al. (2011) utilized an informational masking approach where listeners reported which of six targets (narrowband four-tone sequences that formed clear patterns) occurred amidst multiple narrowband four-tone maskers. Relative to conditions that presented only the masker-plus-target, the largest improvement in performance was produced by having the masker precede the masker-plus-target ipsilaterally (as in EE paradigms). Lesser but still considerable improvement was also produced by presenting the preceding masker in the contralateral ear. Thus, "previewing" the masker reduced its informational masking upon target

[1] Throughout this report, central processing and mechanisms refer to neural processing in the auditory brainstem and beyond. Influences of higher-level central processes such as attention and cognition on spectral context effects are reviewed in the General Discussion.

identification (see also Richards et al., 2004), but this trial structure is also consistent with enhancing detection of the target through EEs. Finally, Byrne, Stellmack, and Viemeister (2013) used a subjective pitch salience task to measure EEs. Listeners reported if a salient pitch was perceived when the target component was either present (and enhanced via EEs) or absent. Among other conditions, multitone complex maskers preceded masker-plus-target complexes to produce intensity EEs, or two segments were presented with one component changed in frequency to produce frequency EEs (cf. Erviti et al., 2011). In both cases, pitch salience was higher when stimuli were presented ipsilaterally and smaller but still present when stimuli were presented contralaterally. Altogether, these findings are consistent with central auditory processing contributing to EEs.

Both SCEs and EEs appear to receive contributions from both peripheral and central processing,[2] but the relative contributions of each are unclear. Across these studies, effect magnitudes generally appear to be smaller when central processing is prioritized (contralateral presentation of context and target stimuli) than in conditions that prioritize peripheral processing (ipsilateral presentation). This might be indicative of peripheral processing contributing more to producing these context effects than central processing. Statistical analyses of these decreases in effect magnitude for contralateral versus ipsilateral stimulus presentation would clarify this possibility, but this has not been done consistently in previous studies. Furthermore, wide variation in stimuli and tasks potentially challenge the extent to which strong conclusions can be made about the geneses of these context effects.

The present experiments manipulated ear(s) of presentation in order to elucidate the relative contributions of peripheral and central processing to SCEs and EEs in speech perception. Listeners identified target consonants as "da" or "ga" following context sentences with F_3 frequency regions bandpass or bandstop filtered in order to produce SCEs or EEs, respectively (see Stilp, 2019). In Experiment 1, trial sequences were presented ipsilaterally, bilaterally, or contralaterally (context in one ear followed by target in the opposite ear). In Experiment 2, monotic context effects were compared to dichotic effects where the context sentence and target were presented to one ear while the spectral complement to the context sentence was presented to the opposite ear. This arrangement directly pits peripheral mechanisms (i.e., the ear receiving context and target, which is sufficient to produce the context effect) against central mechanisms (i.e., the two complementary contexts combining centrally to produce a spectrally neutral stimulus that would not bias perception, thus diminishing the context effect). Across experiments, patterns of results are suggestive of both SCEs and EEs being primarily but not exclusively produced by peripheral processing.

2. Experiment 1

2.1. Methods

2.1.1. Listeners

Forty undergraduate students were recruited from the Department of Psychological and Brain Sciences at the University of Louisville (n = 20 in Experiment 1a, n = 20 in Experiment 1b). No listener participated in multiple experiments. Listeners were native English speakers with self-reported normal hearing, and received course credit for their participation. All procedures involving human listeners were approved by the University of Louisville Institutional Review Board.

2.1.2. Stimuli

2.1.2.1. Target consonants. Target consonants were a series of ten morphed natural tokens from /ga/ to /da/ (Stephens and Holt, 2011). F_3 onset frequencies varied from 2338 Hz (/ga/ endpoint) to 2703 Hz (/da/ 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. Categorization of these targets has been shown to be influenced by both SCEs and EEs (Stilp, 2019).

2.1.2.2. Context sentence. 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). Spectral energy was approximately equal (within 1 dB) across the key frequency regions [1700–2700 Hz (low F_3) and 2700–3700 Hz (high F_3)]. This is the same stimulus that elicited SCEs and EEs for the consonant targets in Stilp (2019). The sentence was processed by two types of filters, each applied to each of these frequency regions. A bandpass filter with 20 dB filter gain was implemented to add a spectral peak to the context sentence (in order to produce an SCE). Bandpass filters had 50-Hz transition regions and used 1200 coefficients. A bandstop filter with 1-Hz transition regions and 1000 coefficients was used to remove energy from the selected frequency region (in order to produce an EE). All filters were created using the fir2 function in MATLAB (MathWorks, Inc., Natick, MA). This produced four versions of the context sentence: low-F_3 amplified (via bandpass filter), high-F_3 amplified (via bandpass filter), low-F_3 attenuated (via bandstop filter), and high-F_3 attenuated (via bandstop filter).

All context and target sounds were low-pass filtered at a cutoff frequency of 5000 Hz and 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. Three trial types were generated: ipsilateral presentation of context and target, bilateral, or contralateral (Fig. 1).

2.1.3. Procedure

Stimuli were resampled at 44,100 Hz sampling rate and presented at an average sound pressure level (SPL) of approximately 70 dB via circumaural headphones (Beyer-Dynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY). Listeners participated individually in single-wall sound-isolating booths (Acoustic Systems, Inc., Austin, TX). Following acquisition of informed consent, listeners first completed a brief practice session comprised of 20 trials. Practice trials presented a context sentence from the AzBio corpus (Spahr et al., 2012) followed by one of the two endpoints from the consonant continuum. Listeners responded by clicking the mouse to indicate whether the target consonant sounded more like "da" or "ga." Feedback was provided. Listeners were required to categorize consonants 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 consonants with 80% accuracy after the third practice session, they were not allowed to participate in the main experiment. All listeners passed the practice session.

Listeners then proceeded to the main experiment, which consisted of three blocks: ipsilateral presentation of contexts and targets on each trial, bilateral presentation, or contralateral presentation. Each block utilized the method of constant stimuli. Experiment 1a tested SCEs in all three modes of presentation, and Experiment 1b tested EEs in all three modes of presentation. Each

[2] Contributions of central processing have also been confirmed through spatialization of context and target stimuli, with larger SCE magnitudes (Watkins, 1991; Feng and Oxenham, 2018b) and EE magnitudes (Šerman et al., 2008) for stimuli sharing the same ITD versus different ITDs.

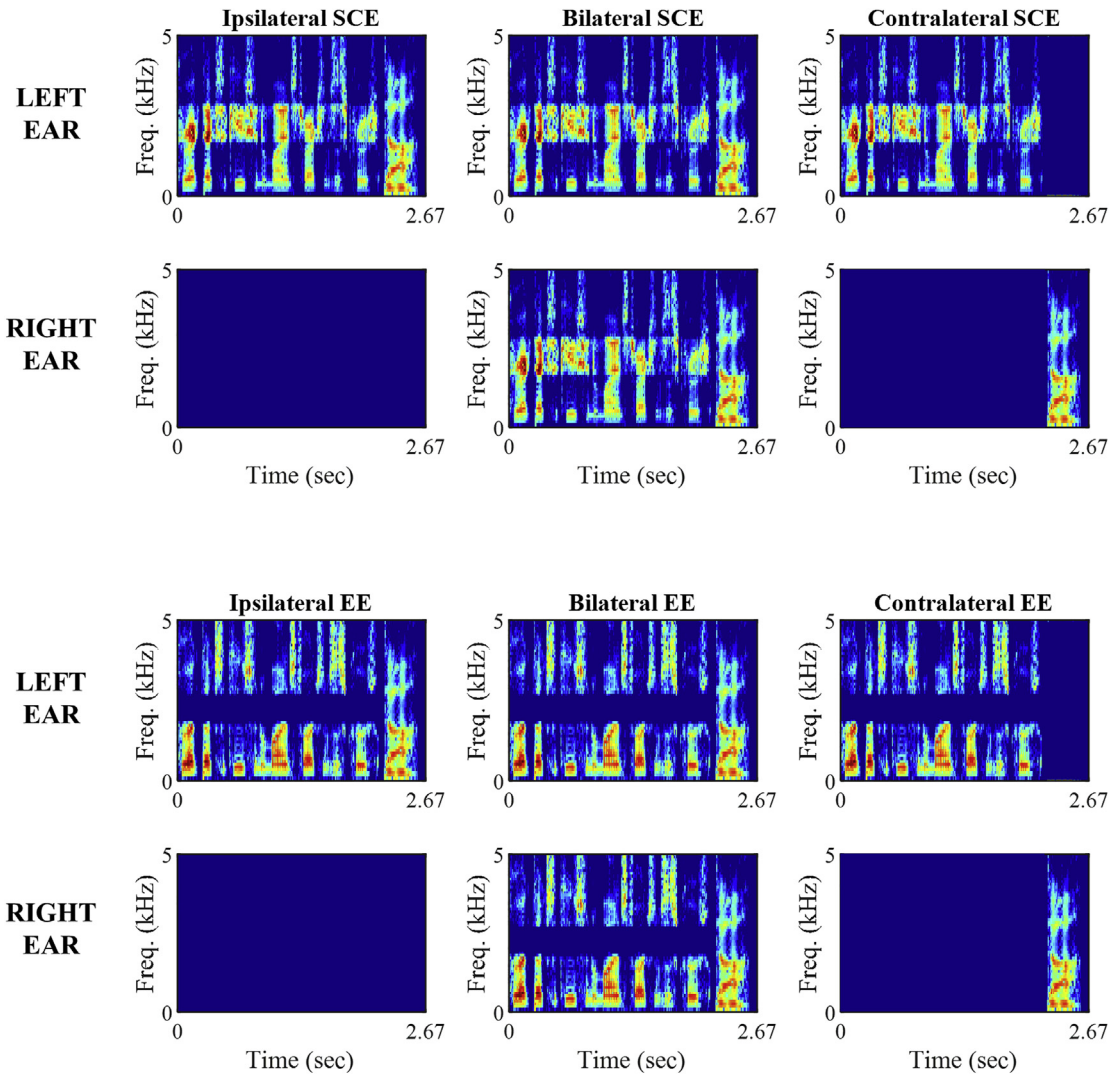

Fig. 1. (color online) Sample trials in Experiment 1. Each pair of spectrograms depicts a trial where the context sentence has a low-F_3 (1700–2700 Hz) spectral peak (top two rows; Experiment 1a) or spectral notch (bottom two rows; Experiment 1b) preceding the /ga/ endpoint target. Figure titles denote the ear(s) of presentation and resulting spectral context effect. Designations of left ear and right ear above are for illustrative purposes; context and target stimuli were counterbalanced across left and right ears in each block. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

block consisted of 160 trials (2 contexts x 10 targets x 8 repetitions), which were presented in random orders. For ipsilateral and contralateral presentations, ear of presentation was balanced across trials in each block. 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.

2.2. Results

A performance criterion was implemented where participants were required to maintain 80% accuracy on consonant endpoint stimuli across the experiment. All 20 listeners in Experiment 1a met this criterion, but one listener failed to meet this criterion in Experiment 1b; his/her data were removed from further analyses.

Responses in each experiment were analyzed using separate mixed-effects logistic models in R (R Development Core Team, 2016) using the lme4 package (Bates et al., 2014). Model architectures were identical across experiments. The dependent variable was modeled as binary ("ga" responses coded as 0 and "da" responses coded as 1). Fixed effects in the model included Target (coded as a continuous variable from 1 to 10 then mean-centered), Frequency (sum coded with the high-F_3 region coded as −0.5 and the low-F_3 region coded as +0.5), Presentation (factor coded with ipsilateral presentation set as the default level), and the interactions between fixed effects. Random slopes were included for each fixed main effect, and a random intercept of participant was included.

Results from Experiment 1a are listed in Table 1 and illustrated in Fig. 2. Listeners responded "da" more often with each rightward step along the target continuum (toward the "da" endpoint; significant effect of Target), and the slopes of the logistic fits to these responses were steeper following bilateral presentation than following ipsilateral presentation (significant Target by Presentation:Binaural interaction). Critically, SCEs occurred (significant main effect of Frequency), and their magnitudes were significantly larger following ipsilateral presentation than following bilateral (significant negative Frequency by Presentation:Bilateral interaction) or contralateral presentations (significant negative Frequency by Presentation:Contralateral interaction). Setting ipsilateral as the default level of

Table 1
Beta estimate (β), standard error (SE), z, and p for the fixed effects of the mixed-effects model for Experiment 1a. Frequency was sum-coded with the level associated with the −0.5 contrast shown in parentheses.

Effect	β	SE	z	p
Intercept	0.066	0.233	0.282	0.778
Target	1.233	0.086	14.367	<0.001
Frequency (High F_3)	3.715	0.227	16.361	<0.001
Presentation:Binaural	0.373	0.268	1.392	0.164
Presentation:Contralateral	−0.371	0.269	−1.379	0.168
Target x Frequency	−0.058	0.076	−0.765	0.444
Target x Presentation:Binaural	0.375	0.068	5.506	<0.001
Target x Presentation:Contralateral	−0.070	0.052	−1.341	0.180
Frequency x Presentation:Binaural	−0.684	0.247	−2.769	0.006
Frequency x Presentation:Contralateral	−3.275	0.208	−15.717	<0.001
Target x Frequency x Presentation:Binaural	0.219	0.124	1.770	0.077
Target x Frequency x Presentation:Contralateral	0.052	0.096	0.541	0.588

Presentation produced contrasts of it against other levels, but not the other levels against each other (e.g., bilateral vs. contralateral). To test that contrast, the model was rerun with contralateral set as the default level of Presentation. SCE magnitudes were significantly smaller following contralateral presentation than following bilateral presentation (Frequency by Presentation:Bilateral interaction: $\beta = 2.590$, s.e. $= 0.215$, $Z = 12.050$, $p < 0.0001$). Finally, SCEs were still significantly greater than zero under contralateral presentation (main effect of Frequency: $\beta = 0.440$, s.e. $= 0.190$, $Z = 2.319$, $p = 0.020$).

SCE magnitudes were calculated following established procedures (Stilp et al., 2015; Stilp, 2019). SCEs were quantified as the number of stimulus steps separating the 50% points on the logistic functions in a given condition (responses following low-F_3-filtered contexts, responses following high-F_3-filtered contexts). For a given presentation mode, since high F_3 is the level of Frequency coded as −0.5, its 50% point is calculated as −Intercept/Slope. The 50% point on the low-F_3 logistic function is calculated as −(Intercept + Frequency)/Slope. The mixed-effect model described above was run with each level of Presentation set as the default level, and SCEs were calculated from the coefficients. Ipsilateral SCEs were largest (3.01 stimulus steps), followed by bilateral SCEs (1.88 stimulus steps), and contralateral SCEs were smallest (0.38 stimulus steps), with all effect magnitudes significantly different from each other (Fig. 2).

Results from Experiment 1b are listed in Table 2 and illustrated in Fig. 2. Listeners responded "ga" more often overall following ipsilateral presentations (significant negative Intercept), but still responded "da" more often with each rightward step along the target continuum (toward the "da" endpoint; significant effect of Target). As in Experiment 1a, the slopes of the logistic fits to these responses were steeper following bilateral presentation than following ipsilateral presentation (significant Target by Presentation:Bilateral interaction). These slopes also varied as a function of context frequency, being shallower for the low-F_3 notched context sentences (negative Target by Frequency interaction). Critically, EEs occurred (significant main effect of Frequency, the negative sign indicating response shifts in the opposite direction of that for SCEs). EE magnitudes were significantly larger following

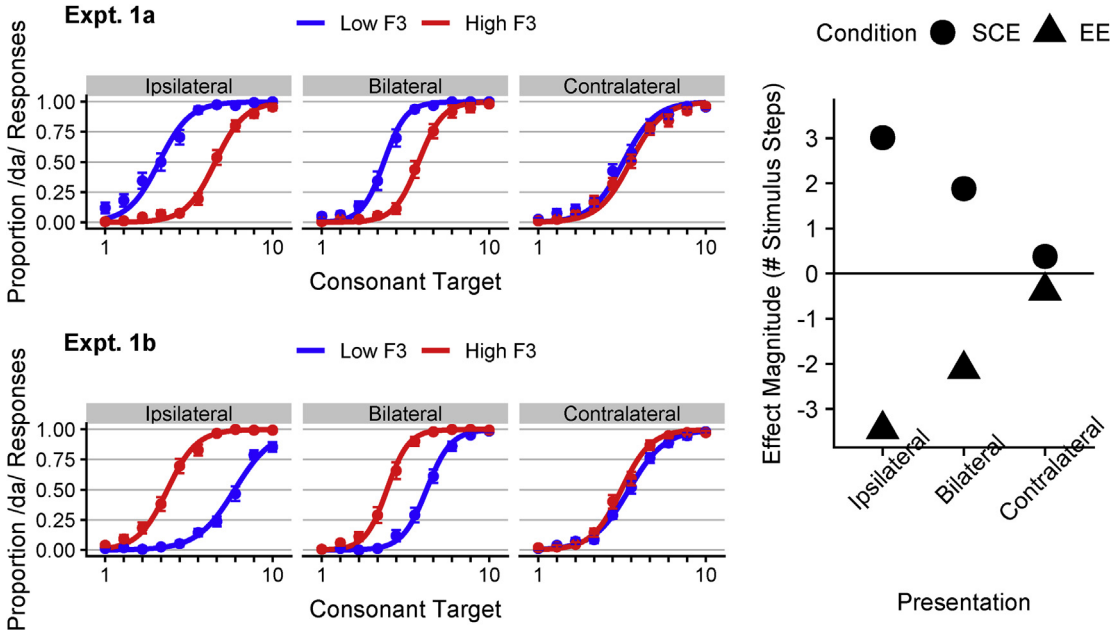

Fig. 2. (color online) Results from Experiment 1. The mean probabilities of listeners responding "da" to each consonant target are plotted in each condition; logistic regressions depict mixed-effects model fits to the data. Error bars depict standard error of the mean. Blue indicates conditions where the context low F_3 frequencies were amplified (Expt. 1a, top left) or attenuated (Expt. 1b, bottom left); red indicates conditions where the context high F_3 frequencies were amplified or attenuated. Context effect magnitudes calculated from the mixed-effects models are plotted at right (see text for details). SCEs are coded as positive shifts and EEs are coded as negative shifts to reflect the fact that these effects influence perception in complementary directions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Beta estimate (β), standard error (SE), z, and p for the fixed effects of the mixed-effects model for Experiment 1b. Frequency was sum-coded with the level associated with the −0.5 contrast shown in parentheses.

Effect	β	SE	z	p
Intercept	−0.479	0.182	−2.634	0.008
Target	1.124	0.064	17.475	<0.001
Frequency (High F_3)	−3.878	0.220	−17.607	<0.001
Presentation:Binaural	0.468	0.279	1.680	0.093
Presentation:Contralateral	0.307	0.183	1.677	0.094
Target x Frequency	−0.284	0.079	−3.622	<0.001
Target x Presentation:Binaural	0.489	0.074	6.565	<0.001
Target x Presentation:Contralateral	0.023	0.056	0.413	0.679
Frequency x Presentation:Binaural	0.456	0.267	1.710	0.087
Frequency x Presentation:Contralateral	3.447	0.219	15.763	<0.001
Target x Frequency x Presentation:Binaural	0.097	0.136	0.715	0.475
Target x Frequency x Presentation:Contralateral	0.075	0.106	0.705	0.481

ipsilateral presentation than following contralateral presentation (significant Frequency by Presentation:Contralateral interaction; positive coefficient indicates less negative, i.e. smaller, effect), and trended in that direction for bilateral presentations (Frequency by Presentation:Bilateral interaction $p = 0.087$). To test for changes in EE magnitudes between bilateral and contralateral presentation modes, the model was rerun with contralateral set as the default level of Presentation. EE magnitudes were significantly smaller following contralateral presentation than following bilateral presentation (Frequency by Presentation:Bilateral interaction: $\beta = -2.991$, s.e. $= 0.232$, $Z = -12.885$, $p < 0.0001$). EEs in contralateral stimulus presentation were still significantly different from zero (main effect of Frequency: $\beta = -0.432$, s.e. $= 0.175$, $Z = -2.464$, $p = 0.014$).

EE magnitudes were calculated using the same procedures detailed above for calculating SCE magnitudes. The mixed-effect model was run with each level of Presentation set as the default level, and EEs were calculated from the coefficients. Ipsilateral EEs (−3.45 stimulus steps) trended toward being significantly larger than diotic EEs (−2.12 stimulus steps), which were significantly larger than contralateral EEs (−0.38 stimulus steps) (Fig. 2).

2.3. Discussion

Across several studies, spectral context effects in auditory perception followed a general pattern where ipsilateral presentations of context and target stimuli produced larger perceptual shifts than contralateral presentations. This pattern was observed in studies of SCEs (Watkins, 1991; Holt and Lotto, 2002; Lotto et al., 2003; Feng and Oxenham, 2018b) and EEs (Erviti et al., 2011; Kidd et al., 2011; Carcagno et al., 2012; Byrne et al., 2013). However, wide variation in the stimuli tested and inconsistency in the statistical analyses of this difference obfuscated the relative contributions of peripheral and central processing to these effects. Experiment 1 utilized a single set of context and target stimuli to produce both SCEs and EEs, providing direct statistical tests of changes in effect magnitudes across presentation modes. Both context effects occurred in contralateral presentation, but the magnitudes were significantly diminished compared to ipsilateral and bilateral presentations. This mirrors the pattern of results outlined in the Introduction, supporting relatively greater contributions from peripheral processing than central processing.

SCEs and EEs have been produced by both shorter-duration (e.g., tens to a few hundred milliseconds) and longer-duration (e.g., one-plus seconds) contexts alike (see Stilp, 2020 for review). In studies that directly varied context duration, both SCEs (Holt, 2006) and EEs (Viemeister, 1980) exhibited larger magnitudes as context duration increased. This result is consistent with the presence of adaptation-related mechanisms throughout the ascending auditory system (i.e., both peripherally and centrally) and the trend toward longer adaptation time constants in higher auditory nuclei. Short-duration contexts are sufficient to produce adaptation in (at least) the auditory periphery, whereas longer-duration contexts also recruit adaptation from relatively higher locations in the auditory system. The present experiments utilized longer-duration (sentence-length) contexts, which presented ample opportunity for slower central adaptation to contribute to these context effects. Yet, while contralateral presentation of contexts and targets did produce SCEs and EEs, their magnitudes were substantially smaller than those observed in ipsilateral presentation. Were the present results produced by using short-duration contexts, it could be argued that central processing was not adequately engaged to test this research question, but this was not the case. This further supports the interpretation that SCEs and EEs are primarily peripheral effects.

Monotic context effects were significantly larger than effects in bilateral presentations (SCEs) or trended in that direction (EEs). To the best of our knowledge, no previous published studies tested both of these presentation modes in the same participants. Instead, previous investigations generally utilized one or the other to compare against effects of contralateral stimulus presentation. One possible explanation for this pattern of results is contralateral inhibition elicited by the medial olivocochlear (MOC) bundle. Efferent projections from the MOC terminating on the outer hair cells of the contralateral cochlea can modulate cochlear gain (Guinan, 2018). As such, stimulation in the contralateral ear can have an inhibitory effect on responses in the ipsilateral ear, as would occur under bilateral stimulus presentation. Further research is needed to clarify this possibility for both SCEs and EEs.

The context stimuli that produce context effects in Experiment 1 are spectrally complementary to each other: the stopbands that produce EEs span the same frequency ranges as the passbands that produce SCEs (Fig. 1). This complementarity allows for a novel test of the relative contributions of peripheral versus central processing to these spectral context effects. Stopband and passband versions of the context sentence bias perception individually, but combining the two would create a context that is spectrally neutral and thus not produce an EE nor an SCE. Experiment 2 utilized this approach to explore the degree to which the central auditory system could diminish a context effect produced in the auditory periphery. During the ipsilateral presentation of the context stimulus in one ear (e.g., stopband-filtered sentence) before the target, the spectrally complementary context stimulus (e.g., passband-filtered version of that sentence) was presented concurrently in the contralateral ear. Context effects in these dichotic trials were compared to those observed in monotic trials, which now serve as

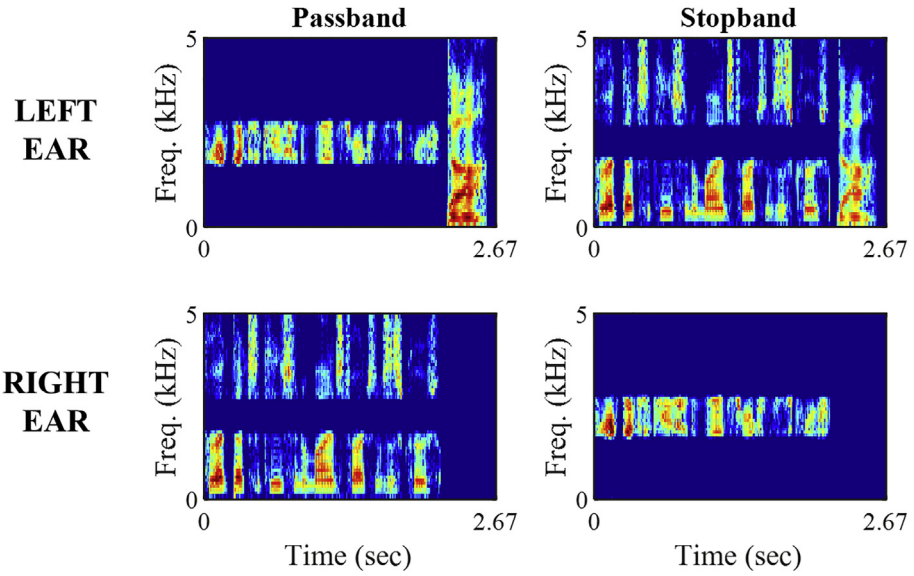

Fig. 3. (color online) Sample trials in Experiment 2. Each pair of spectrograms depicts a trial where context sentences have low-F_3 (1700–2700 Hz) frequencies processed by the passband or stopband filter preceding the /ga/ endpoint target. Trials are named for the filter applied to the target-ipsilateral context. Designations of left ear and right ear above are for illustrative purposes; context and target stimuli were counterbalanced across left and right ears in each block. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the baseline. Substantially diminished context effect magnitudes in dichotic trials would support considerable contribution by central auditory processing to these effects. Conversely, highly comparable effect magnitudes across dichotic and monotic trials would indicate minimal contribution from the central auditory system, as effects were primarily resolved in the auditory periphery and unperturbed by addition of the spectrally complementary context in the central auditory system.

3. Experiment 2

3.1. Methods

3.1.1. Listeners

Twenty undergraduate students were recruited from the Department of Psychological and Brain Sciences at the University of Louisville. No listener participated in Experiment 1a or 1b. Listeners were native English speakers with self-reported normal hearing, and received course credit for their participation.

3.1.2. Stimuli

Experiment 2 utilized the same target consonant stimuli as Experiment 1. As for context sentences, the same bandstop filters from Experiment 1 were used in Experiment 2. These bandstop filters were then inverted (bandstop frequencies became bandpass frequencies, bandpass frequencies became bandstop frequencies) to create spectrally complementary bandpass filters for the low-F_3 and high-F_3 frequency regions. These bandpass and bandstop filters both had 1-Hz transition regions and 1000 coefficients. Filters were again created using the fir2 function in MATLAB (MathWorks, Inc., Natick, MA).

In two conditions of Experiment 2, a context and target were presented to one ear while the spectrally complementary context was presented to the opposite ear (Fig. 3). In one block, a passband context (which produced SCEs in Stilp, 2019) was presented ipsilateral to the target stimulus (/da/-/ga/ continuum member). Concurrent to the passband context, the complementary stopband context (which produced EEs in Stilp, 2019 and Experiment 1) was presented in the opposite ear without any subsequent target stimulus (Fig. 3). These trials are labeled as passband trials because that describes the filtering done to the context ipsilateral to the target stimulus. Similarly, in the trials designated as stopband, a stopband context was presented ipsilateral to the target stimulus while the corresponding passband context was presented to the opposite ear without any target stimulus.

When creating these trials, full-spectrum context sentences were set to 70 dB SPL before stopband or passband filtering. Stimulus level was not reset to 70 dB following filtering (as was done in Experiment 1) but left at their native levels (average levels of 57.37 dB SPL for passband contexts and 69.70 dB SPL for stopband contexts). This was done so that when complementary contexts were combined in the central auditory system, the full-spectrum context would approximately match the spectrally neutral seed stimulus.[3] Finally, monotic SCE trials and monotic EEs trials were included as baseline conditions. Presentation levels for monotic contexts matched those presented in dichotic trials.

3.1.3. Procedure

Experiment 2 followed the same procedures as Experiment 1 but now with four experimental blocks (monotic bandpass, dichotic bandpass, monotic bandstop, dichotic bandstop). As in Experiment 1, ear of presentation was balanced across trials within each block. The entire session lasted approximately 45 min.

3.2. Results

The same performance criterion requiring 80% accuracy on consonant endpoint stimuli was employed. All listeners met this criterion, thus all data were included in analyses. Results were analyzed in two separate mixed-effects models with identical architectures (one model analyzing results from passband trials, the

[3] A separate version of these stimuli were created with all filtered context stimuli set to 70 dB SPL. A separate group of participants was tested using these stimuli, and the patterns of results replicated those reported below. Thus, differences in passband context level and stopband context level were not a confound.

other analyzing results from stopband trials). In each model, the dependent variable was again binary ("ga" responses coded as 0 and "da" responses coded as 1), and fixed effects included Target (mean-centered continuous variable), Frequency (high-F_3 region coded as 0, low-F_3 region coded as 1), Presentation (monotic coded as 0, dichotic coded as 1), and the interactions between fixed effects. Categorical coding was employed to test for effects of presentation mode while simultaneously estimating context effect magnitudes. Random slopes were included for each fixed main effect, and a random intercept of participant was included.

Results from the passband model are listed in Table 3 and illustrated in Fig. 4. Listeners responded "da" more often with each rightward step along the target continuum (significant effect of Target). Listeners' "da" responses significantly increased when the passband frequency was changed from high-F_3 to low-F_3 (consistent with SCEs, significant positive effect of Frequency). Critically, the negative Frequency by Presentation interaction indicates that SCE magnitudes were smaller in dichotic trials than in monotic trials. Monotic SCEs were 3.05 stimulus steps and dichotic SCEs were 2.54 stimulus steps, calculated following the same methods described in Experiment 1.

Results from the stopband model are listed in Table 4 and also illustrated in Fig. 4. Listeners again responded "da" more often with each rightward step along the target continuum (significant effect of Target). Listeners' "da" responses significantly decreased when the stopband frequency was changed from high-F_3 to low-F_3 (consistent with EEs, significant negative effect of Frequency). The positive Frequency by Presentation interaction approached statistical significance, indicating that EE magnitudes were modestly smaller in dichotic trials than in monotic trials. Monotic EEs were −3.04 stimulus steps and dichotic EEs were −2.72 stimulus steps, calculated following the same methods described in Experiment 1.

3.3. Discussion

While Experiment 1 examined peripheral and central contributions to spectral context effects separately, Experiment 2 put them in direct competition with each other. Monotic presentations of context and target stimuli produced relatively large EEs and SCEs, as also reported in Experiment 1, but the primary question was the extent to which stimuli presented to the contralateral ear could diminish these effects. Central processing, as indexed by the combination of spectrally complementary contexts across both ears, was minimally effective in extinguishing spectral context effects. While SCE magnitudes did diminish in dichotic trials and EE magnitudes trended in that direction, effect magnitudes were still relatively large overall. Consistent with the results of Experiment 1, these results are suggestive of both spectral context effects exhibiting a primarily but not exclusively peripheral genesis.

Table 3
Beta estimate (β), standard error (SE), z, and p for the fixed effects of the mixed-effects model for passband trials of Experiment 2. Frequency and Presentation mode were categorically coded with the default levels (coded as 0 in the model) shown in parentheses.

Effect	β	SE	z	p
Intercept	−1.218	0.335	−3.630	<0.001
Target	1.029	0.077	13.344	<0.001
Frequency (High F_3)	3.143	0.266	11.813	<0.001
Presentation (Monotic)	0.059	0.263	0.225	0.822
Target x Frequency	0.027	0.062	0.442	0.658
Target x Presentation	0.019	0.059	0.327	0.744
Frequency x Presentation	−0.481	0.194	−2.479	0.013
Target x Frequency x Presentation	−0.004	0.083	−0.052	0.959

Experiment 2 extends the results reported by Feng and Oxenham (2018b) in their investigation of SCEs in vowel categorization. They presented context sentences that were processed by spectral envelope difference filters, which reshaped the context spectrum in order to emphasize the difference between target item spectra (Watkins, 1991). The context sentence presented to one ear was processed by one difference filter (i.e., the spectrum of /ɪ/ minus the spectrum of /ɛ/) while the context presented to the other ear was processed by the other difference filter (/ɛ/ minus /ɪ/). Similar to Experiment 2, the vowel target was presented to only one ear. However, their approach did not provide a strong test of whether peripheral context effects could cancel centrally for several reasons. First, a different sentence was presented to each ear, which eliminates the possibility of spectra completely cancelling when combined in the central auditory system. The present method of passband and stopband filtering a single sentence produced a centrally combined stimulus that is more spectrally neutral. Second, while Feng and Oxenham's method explored the presence or absence of SCEs, Experiment 2 expanded this test to the occurrences and predicted cancellations of SCEs and EEs, using context effects in complementary directions to test the present hypotheses. Third, their effect magnitudes were much smaller than those in the present experiment (for target stimuli generated using the Praat method,[4] mean shifts of 0.84 steps for ipsilateral presentation of context and target stimuli in their Experiment 1a and 0.79 steps for dichotic presentation in their Experiment 2a). Smaller effect magnitudes increase the difficulty of distinguishing effect magnitudes from each other (and from zero). Finally, their listeners were instructed to attend to one of the two competing sentences, which could limit the ability of spectra to combine and cancel centrally; these manipulations of listeners' attention (as well as similar work by Bosker et al., 2019) are discussed further in the General Discussion. Listeners in the present experiment were not given any explicit directions regarding attention. Even with all of these considerations, the combined contexts in the central auditory system had an overall weak influence on diminishing context effects produced at the periphery (Fig. 4).

Monaural and binaural processing of Experiment 2 stimuli merit additional consideration. Dichotic trials were designed to capitalize on the binaural integration of projections from cochlear nuclei to the superior olivary complex. But, there also exist monaural projections from cochlear nuclei directly to the inferior colliculus, forming a central monaural pathway. It is an open question whether monaural stimulus representations in the inferior colliculus (originating in the ear that received both context and target, producing the context effect) could reduce the influence of the binaural stimulus representation (where spectrally complementary contexts combined via binaural integration), as ultimately both pathways feed forward toward perceptual decisions and responses. An intriguing direction for future research is to collect physiological recordings in the inferior colliculus to compare the relative prominence of monaural and binaural stimulus representations in this paradigm. However, following previous studies (Watkins, 1991; Holt and Lotto, 2002; Lotto et al., 2003; Erviti et al., 2011; Kidd et al.,

[4] Feng and Oxenham (2018b) also tested target vowels generated using the 'Watkins method', which produces a stimulus continuum through proportionate blending of the spectral envelopes of the endpoint stimuli. This method is not recommended as it produces mid-continuum stimuli that are physically impossible to produce (i.e., containing spectral peaks corresponding to the lower F_1 frequency of /ɪ/ and the higher F_1 frequency of /ɛ/ simultaneously). The Praat method is preferred as it is more realistic (i.e., mid-continuum stimuli possessing a single spectral peak at an F_1 frequency intermediate to those found in /ɪ/ and /ɛ/). While the present stimuli were not generated using the Praat method per se, they are analogous to its products (mid-continuum stimuli possessing an F_3 transition onset frequency that is intermediate to those found in /d/ and /g/).

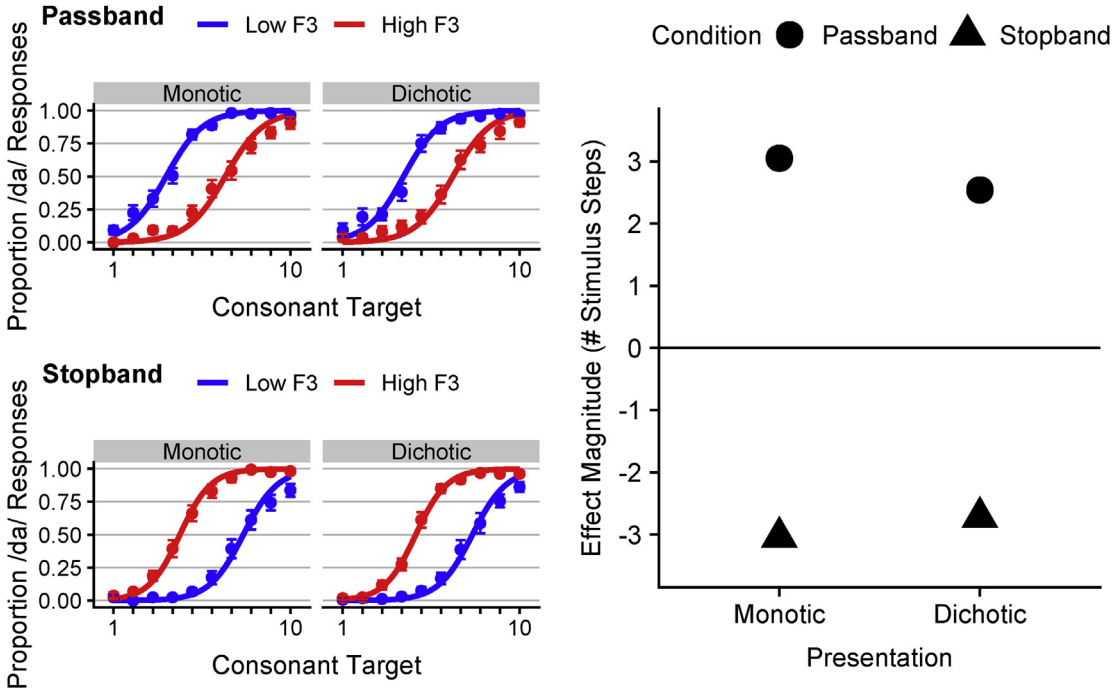

Fig. 4. (color online) Results from Experiment 2. The mean probabilities of listeners responding "da" to each consonant target are plotted in each condition; logistic regressions depict mixed-effects model fits to the data. Error bars depict standard error of the mean. Blue indicates conditions where the context low F_3 frequencies were processed by bandpass (Expt. 1a, top left) or bandstop filters (Expt. 1b, bottom left); red indicates conditions where the context high F_3 frequencies were bandpass or bandstop filtered. Context effect magnitudes calculated from the mixed-effects models are plotted at right (see text for details). SCEs are coded as positive shifts and EEs are coded as negative shifts to reflect the fact that these effects influence perception in complementary directions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 4
Beta estimate (β), standard error (SE), z, and p for the fixed effects of the mixed-effects model for stopband trials of Experiment 2. Frequency and Presentation mode were categorically coded with the default levels (coded as 0 in the model) shown in parentheses.

Effect	β	SE	z	p
Intercept	1.425	0.287	4.972	<0.001
Target	1.266	0.090	14.048	<0.001
Frequency (High F_3)	−3.852	0.344	−11.204	<0.001
Presentation (Monotic)	−0.438	0.224	−1.955	0.051
Target x Frequency	−0.113	0.073	−1.562	0.118
Target x Presentation	−0.009	0.073	−0.119	0.905
Frequency x Presentation	0.429	0.224	1.918	0.055
Target x Frequency x Presentation	0.016	0.099	0.159	0.873

2011; Carcagno et al., 2012; Byrne et al., 2013), it is more appropriate to view the geneses of these effects as being peripheral/central rather than monaural central/binaural central. The neural mechanisms hypothesized to underlie each effect (SCEs: simple neural adaptation; EEs: adaptation of inhibition) are available at multiple sites throughout the ascending auditory system, starting in the auditory periphery. EEs (and likely SCEs) are thought to occur and repeat at multiple successive levels of the auditory system (Nelson and Young, 2010; Carcagno et al., 2012; Feng and Oxenham, 2015; Feng et al., 2018). Therefore, the most parsimonious perspective is for effects to begin (and from the results of the present experiments, exert relatively greater influence) in the auditory periphery before recurring (with relatively less influence) in the central auditory system.

4. General Discussion

Auditory enhancement effects (EEs) and spectral contrast effects (SCEs) have both been widely reported in studies of auditory perception. These spectral context effects are thought to be produced by mechanisms related to neural adaptation (adaptation of inhibition and simple neural adaptation, respectively). These effects appear to receive contributions from both peripheral (cochlear and including the auditory nerve) and central (cochlear nucleus and beyond) processing, but the relative contributions of each level of processing are less clear. The present experiments addressed this question through manipulations of ear(s) of presentation while measuring EEs and SCEs in speech perception.

In Experiment 1, the filtered context sentence and the subsequent consonant target were presented ipsilaterally, bilaterally, or contralaterally on each trial. Ipsilateral (a proxy for measuring contributions from peripheral processing) and bilateral presentations produced substantially larger context effects than contralateral stimulus presentation (a proxy for measuring contributions from central processing). While all three presentation modes affected speech perception, this pattern of results (which was observed for EEs and SCEs alike) is suggestive of peripheral processing contributing more to these context effects than central processing. In Experiment 2, peripheral and central contributions were pitted directly against one another. Monotic presentation of context and target stimuli, already shown to produce large context effects in Experiment 1, served as the baseline condition. Dichotic trials added the spectrally complementary context sentence to the opposite ear, which was predicted to combine with the target-ipsilateral context centrally and diminish the context effect. This contralateral stimulus had a very weak effect on performance; relative to the monotic condition, SCE magnitudes decreased but were still large overall, and EE magnitudes were not significantly diminished. These results are also consistent with a primarily peripheral locus for the genesis of these spectral context effects.

However, the present results indicate that the central auditory system also contributes to these context effects. Contralateral presentations of context and target stimuli in Experiment 1 still produced context effects (Fig. 2), albeit with smaller magnitudes than other presentation modes. Addition of a spectrally complementary context stimulus did slightly but significantly reduce SCE magnitudes in Experiment 2 (Fig. 4). Other investigations have demonstrated contributions of central processing by spatializing the context and target stimuli. Context effect magnitudes were smaller when stimuli had different ITDs compared to the same ITDs (SCEs: Watkins, 1991; Feng and Oxenham, 2018b; EEs: Šerman, Semal and Demany, 2008). Finally, one cannot unambiguously identify underlying neural mechanisms through behavioral data alone. Direct physiological recordings are necessary in order to definitively establish the relative contributions of neural processing at different stages of the auditory system toward context effects in auditory perception.

Conclusions drawn from the present results must be viewed in the light of the methodological decisions made for these experiments. First, the present experiments utilized speech stimuli high in spectrotemporal variability as both context and target stimuli. This is not uncommon in investigations of SCEs, but investigations of EEs typically employ spectrotemporally simpler stimuli such as filtered noise or multitone complexes. Second, categorization of complex sounds is the task typically employed to measure SCEs, but a wide variety of tasks can be used to measure EEs (e.g., detection, pitch salience rating, pitch movement judgments, pattern recognition, vowel recognition, simultaneous masking, forward masking; see Introduction). Trials in the present experiments most closely resembled simultaneous masking paradigms for measuring EEs, whose results do not always pattern the same way as forward masking paradigms (Kreft and Oxenham, 2017). Third, the context preceded the target in each trial in what are termed forwards context effects. Both SCEs and EEs have been reported using backwards context effect paradigms (target precedes the context), but often with smaller magnitudes than forwards paradigms (see Stilp, 2020 for a review of backwards SCEs and Byrne et al., 2013 for an example of backwards EEs). Fourth, SCEs and EEs have both been reported following shorter-duration contexts as well as longer-duration contexts. The context sentence tested here (2200 ms) is of relatively longer duration compared to other contexts tested in the literature. Multi-second contexts would be expected to recruit central auditory processing given its longer time constants of adaptation. That fact makes it all the more surprising that central influences on speech categorization were as small as they were here. Finally, the ISI between context and target stimuli on each trial was brief (50 ms). This might favor the shorter time constants of neural adaptation in the periphery, but both effects have been observed with ISIs between contexts and targets upward of 5 s or longer (SCEs: Broadbent and Ladefoged, 1960; EEs: Viemeister, 1980). Still, statistical analyses of context effect magnitudes in the present experiments produced parallel patterns of results to a wide variety of studies using diverse tasks and stimuli listed in the Introduction. This symmetry bolsters the generalizability of the present findings to EE and SCE research at large.

The medial olivocochlear reflex (MOCR), cited as a potential explanation for why ipsilateral effects were larger than bilateral effects in Experiment 1, is a potential contributor to both SCEs and EEs. On-frequency stimulation reduces cochlear gain at that frequency upon introduction of subsequent (target) stimuli (Strickland, 2004, 2008; Jennings et al., 2011). For SCEs, this could (at least partially) explain the relatively reduced responsiveness to target frequencies that also occurred in the preceding context, making responses to neighboring frequencies more prominent. For EEs, the spectral notch in the preceding context would not reduce cochlear gain at those frequencies in the target sound, facilitating their detection. In studies that varied notch depth, progressively shallower spectral notch depths resulted in progressively smaller EEs, which may be due to decreased cochlear gain (Summerfield et al., 1987; Stilp, 2019). While an appealing explanation, cochlear implant users have demonstrated both SCEs (Feng and Oxenham, 2018a) and EEs (Wang et al., 2012) without contributions from the MOCR. Thus, the MOCR is not a necessary mechanism for producing these context effects, but this does not rule out its potential contribution to the present results with normal-hearing listeners (and their ostensibly fully functioning MOCRs). Both SCEs and EEs have multiple sources, coarsely classified here as peripheral and central. It is possible that the MOCR is one of several sources that produces these spectral context effects, but it absence (as in cochlear implant users) does not extinguish the effects altogether. Recently, Beim and colleagues (2015) used stimulus-frequency otoacoustic emissions (SFOAEs) to test the role of the MOC system in the production of EEs. SFOAEs exhibited no changes consistent with enhancement, leading the authors to conclude that the MOCR does not play a significant role in EEs. Future research should utilize a variety of experimental paradigms to elucidate potential contributions of the MOCR to these effects.

In this report, peripheral and central mechanisms underlying SCEs and EEs have both been described as being fairly low-level and related to neural adaptation. In actuality, these central mechanisms are not confined to lower-level neural activity but can include higher-level contributions including but not limited to streaming, cognition, and attention. Research on these higher-level contributions to spectral context effects have produced mixed results. Given that research on SCEs has been primarily conducted using speech categorization tasks, this has afforded the opportunity to study interactions between low-level neural processing (i.e., peripheral contributions to SCEs) and higher-level linguistic factors. A series of behavioral studies by Sjerps and colleagues have demonstrated that SCEs appear to occur before these linguistic factors contribute to perception. Acoustically ambiguous evidence for a speech sound is more likely to be perceived as the option which forms a valid word rather than a non-word (Ganong, 1980), but SCEs influence speech categorization before this lexical influence takes place (Sjerps and Reinisch, 2015). SCEs occur similarly across listeners' native and non-native languages (Sjerps and Smiljaniç, 2013; but see Kang et al., 2016). Finally, SCE magnitudes were not modulated by cognitive load incurred by a visual search task in a dual-task paradigm (Bosker et al., 2017). Other research has demonstrated the sensitivity of SCEs to various higher-level influences. Visual information about a talker's gender can shift the location of a boundary between two vowel categories (Glidden and Assmann, 2004), as can expectations about hearing a man or woman speak (Johnson, Strand, & D'Imperio, 1999). Additionally, diminished context effects in contralateral context-target presentations relative to ipsilateral presentations have been viewed as a failure to perceptually group the stimuli together (Summerfield et al., 1987; Kidd et al., 2011).

Two recent studies reported effects of attention on SCEs in vowel categorization. As discussed above, Feng and Oxenham (2018b) presented different context sentences to each ear and the target to only one ear (their Experiment 2A). SCEs were driven by the context sentence presented ipsilaterally to the target sound, and effect magnitudes were reduced (but still present) when listeners were instructed to attend to the contralateral context sentence. When repeating the experiment with target sound presented diotically (their Experiment 2C), SCEs occurred but at a fraction of the magnitude observed in the monotic-target study. Thus, attention overrode the fact that different contexts in each ear biased perception of the target in opposite directions. Similarly, Bosker

et al. (2019) presented target vowels diotically with different context sentences presented to each ear – one sentence spoken by the same talker who produced the target, the other sentence produced by different talkers. When listeners were instructed to attend to the talker who produced the target items, spectral characteristics of the unattended context sentence had no effect on SCE magnitude. When listeners were instructed to attend to the other talkers who did not produce the target items, SCEs were extinguished irrespective of spectral characteristics of attended or unattended talkers. Trial-to-trial variability in talker acoustics has been shown to attenuate SCEs (Assgari and Stilp, 2015; Assgari et al., 2019), but here an added influence of attention contributed to differential results across experiments. Across these two studies, effects of attention were present but relatively weak compared to lower-level acoustic factors. Synthesizing these results with the present studies, it appears that peripheral mechanisms are primarily responsible for producing spectral context effects in speech perception, with far weaker contributions from central processes; whether the contributions of central neural mechanisms and higher-level factors such as attention can be distinguished from each other remains a point of future investigations.

5. Summary

The present experiments sought to illuminate the relative contributions of peripheral and central processing to spectral context effects (SCEs, EEs) in speech perception. Experiment 1 produced substantial decreases in context effect magnitudes for contralateral presentation of contexts and targets as compared to ipsilateral and bilateral presentations. Experiment 2 demonstrated that the magnitudes of peripheral effects (monotic presentation) were minimally affected by presentation of spectrally complementary contexts in the opposite ear (dichotic presentation). Across experiments, results are strongly indicative of primarily although not exclusively peripheral contributions toward SCEs and EEs. Additionally, the parallel patterns of results for SCEs and EEs in both experiments further deepens the direct relationship recently reported between these context effects (Stilp, 2019).

Open practices statement

All data and analysis scripts are available at https://osf.io/kjxdf/.

CRediT authorship contribution statement

Christian E. Stilp: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing - original draft.

Acknowledgements

The author thanks two anonymous reviewers for their feedback and suggestions, and thanks Scott Barrett, Rebecca Davis, Emily Dickey, Ella Beilman, Joshua Lanning, Pratistha Thapa, and Sara Wardrip for assistance with data collection.

References

Assgari, A.A., Stilp, C.E., 2015. Talker information influences spectral contrast effects in speech categorization. J. Acoust. Soc. Am. 138 (5), 3023–3032.
Assgari, A.A., Theodore, R.M., Stilp, C.E., 2019. Variability in talkers' fundamental frequencies shapes context effects in speech perception. J. Acoust. Soc. Am. 145 (3), 1443–1454.
Bates, D.M., Maechler, M., Bolker, B., Walker, S., 2014. lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from. http://cran.r-project.org/package=lme4.
Beim, J.A., Elliott, M., Oxenham, A.J., Wojtczak, M., 2015. Stimulus frequency otoacoustic emissions provide no evidence for the role of efferents in the enhancement effect. J. Assoc. Res. Otolaryngol. 16, 613–629.
Bosker, H.R., Reinisch, E., Sjerps, M.J., 2017. Cognitive load makes speech sound fast, but does not modulate acoustic context effects. J. Mem. Lang. 94, 166–176.
Bosker, H.R., Sjerps, M.J., Reinisch, E., 2019. Spectral contrast effects are modulated by selective attention in "cocktail party" settings. Atten. Percept. Psychophys. 1–15.
Broadbent, D.E., Ladefoged, P., 1960. Vowel judgements and adaptation level. Proc. Biol. Sci. 151, 384–399.
Byrne, A.J., Stellmack, M.A., 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., Viemeister, N.F., 2013. The salience of enhanced components within inharmonic complexes. J. Acoust. Soc. Am. 134 (4), 2631–2634.
Carcagno, S., Semal, C., Demany, L., 2012. Auditory enhancement of increments in spectral amplitude stems from more than one source. J. Assoc. Res. Otorhinolaryngol. 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., 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: Hardcastle, W.J., Laver, J. (Eds.), The Handbook of Phonetic Sciences. Blackwell Publishing Ltd, Oxford, pp. 507–538.
Delgutte, B., Hammond, B.M., Kalluri, S., Litvak, L.M., Cariani, P.A., 1996. Neural encoding of temporal envelope and temporal interactions in speech. In: Ainsworth, W., Greenberg, S. (Eds.), Proceedings of Auditory Basis of Speech Perception. European Speech Communication Association, pp. 1–9.
Erviti, M., Semal, C., Demany, L., 2011. Enhancing a tone by shifting its frequency or intensity. J. Acoust. Soc. Am. 129 (6), 3837–3845.
Feng, L., Mehta, A.H., Oxenham, A.J., 2018. Neural correlates of auditory enhancement in humans. BioRxiv 1–23. https://doi.org/10.1101/458521.
Feng, L., Oxenham, A.J., 2015. New perspectives on the measurement and time course of auditory enhancement. J. Exp. Psychol. Hum. Percept. Perform. 41 (6), 1696–1708.
Feng, L., 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., Oxenham, A.J., 2018b. Spectral contrast effects produced by competing speech contexts. J. Exp. Psychol. Hum. Percept. Perform. 44 (9), 1447–1457.
Ganong, W.F., 1980. Phonetic categorization in auditory word perception. J. Exp. Psychol. Hum. Percept. Perform. (1), 110–125.
Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., 1990. "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM." *NIST Order No. PB91-505065*. National Institute of Standards and Technology, Gaithersburg, MD.
Glidden, C.M., Assmann, P.F., 2004. Effects of visual gender and frequency shifts on vowel category judgments. Acoust Res. Lett. Online 5 (4), 132–138.
Guinan, J.J., 2018. Olivocochlear efferents: their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. Hear. Res. 362, 38–47.
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., 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., Kluender, K.R., 2000. Neighboring spectral content influences vowel identification. J. Acoust. Soc. Am. 108 (2), 710–722.
Johnson, K., Strand, E.A., D'Imperio, M., 1999. Auditory-visual integration of talker gender in vowel perception. J. Phonetics 27 (4), 359–384.
Kang, S., Johnson, K., Finley, G., 2016. Effects of native language on compensation for coarticulation. Speech Commun. 77, 84–100.
Kidd, G., Richards, V.M., Streeter, T., Mason, C.R., Huang, R., 2011. Contextual effects in the identification of nonspeech auditory patterns. J. Acoust. Soc. Am. 130 (6), 3926–3938.
Kingston, J., Kawahara, S., Chambless, D., Key, M., Mash, D., Watsky, S., 2014. Context effects as auditory contrast. Atten. Percept. Psychophys. 76, 1437–1464.
Kreft, H.A., Oxenham, A.J., 2017. Auditory enhancement in cochlear-implant users under simultaneous and forward masking. J. Assoc. Res. Otolaryngol. 18 (3), 483–493.
Ladefoged, P., Broadbent, D.E., 1957. Information conveyed by vowels. J. Acoust. Soc. Am. 29 (1), 98–104.
Lanning, J.M., Stilp, C.E., 2020. Natural music context biases musical instrument categorization. Atten. Percept. Psychophys. 1–6. https://doi.org/10.3758/s13414-020-01980-w.
Lotto, A.J., Sullivan, S.C., Holt, L.L., 2003. Central locus for nonspeech context effects on phonetic identification (L). J. Acoust. Soc. Am. 113 (1), 53–56.
Nelson, P.C., 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., 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. R Foundation for Statistical Computing, " Vienna, Austria. Retrieved from. http://www.r-project.org/.
Richards, V.M., Huang, R., Kidd, G., 2004. Masker-first advantage for cues in

informational masking. J. Acoust. Soc. Am. 116 (4), 2278–2288, 1.
Schouten, J., 1940. The residue and the mechanism of hearing. Proc. Koninklijke Nederl. Akademie Wetenschappen 43, 991–999.
Scutt, M.J., Palmer, A.R., 1998. Physiological Enhancement in Cochlear Nucleus Using Single Tone Precursors. In Assoc Res Otolaryngol Abs, 188(A).
Šerman, M., Semal, C., Demany, L., 2008. Enhancement, adaptation, and the binaural system. J. Acoust. Soc. Am. 123 (6), 4412–4420.
Sjerps, M.J., Fox, N.P., Johnson, K., Chang, E.F., 2019. Speaker-normalized sound representations in the human auditory cortex. Nat. Commun. 10, 1–9.
Sjerps, M.J., Reinisch, E., 2015. Divide and conquer: how perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception. J. Exp. Psychol. Hum. Percept. Perform. 41 (3), 710–722.
Sjerps, M.J., Smiljaniç, R., 2013. Compensation for vocal tract characteristics across native and non-native languages. J. Phonetics 41 (3–4), 145–155.
Sjerps, M.J., Zhang, C., Peng, G., 2018. Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context. J. Exp. Psychol. Hum. Percept. Perform. 44 (6), 914–924.
Spahr, A.J., Dorman, M.F., Litvak, L.M., Van Wie, S., Gifford, R.H., Loizou, P.C., et al., 2012. Development and validation of the AzBio sentence lists. Ear Hear. 33 (1), 112–117.
Stephens, J.D.W., 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., 2019. Auditory enhancement and spectral contrast effects in speech perception. J. Acoust. Soc. Am. 146 (2), 1503–1517.
Stilp, C.E., 2020. Acoustic context effects in speech perception. WIREs Cogn Sci. 11 (1), 1–18. https://onlinelibrary.wiley.com/doi/full/10.1002/wcs.1517. (Accessed 14 May 2020).
Stilp, C.E., Alexander, J.M., Kiefte, M., 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., Winn, M.B., 2015. Predicting contrast effects following reliable spectral properties in speech perception. J. Acoust. Soc. Am. 137 (6), 3466–3476.
Summerfield, Q., Haggard, M., Foster, J., Gray, S., 1984. Perceiving vowels from uniform spectra - phonetic exploration of an auditory aftereffect. Percept. Psychophys. 35 (3), 203–213.
Summerfield, Q., Sidwell, A., Nelson, T., 1987. Auditory enhancement of changes in spectral amplitude. J. Acoust. Soc. Am. 81 (3), 700–708.
Viemeister, N.F., 1980. Adaptation of masking. In: Brink, G.v. d., Bilsen, F.A. (Eds.), Psychophysical, Physiological and Behavioural Studies in Hearing. University Press, Delft, pp. 190–198.
Viemeister, N.F., Bacon, S.P., 1982. Forward masking by enhanced components in harmonic complexes. J. Acoust. Soc. Am. 71 (6), 1502–1507.
Wang, N., Kreft, H., Oxenham, A.J., 2012. Vowel enhancement effects in cochlear-implant users. J. Acoust. Soc. Am. 131 (6), EL421–EL426.
Wang, N., 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.