


The influence of phoneme contexts on adaptation in vowel-evoked envelope following responses

Vijayalakshmi Easwar^{1,2}  | Lauren Chung^{1,2}

¹Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA

²Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence

Vijayalakshmi Easwar, National Acoustic Laboratories, Level 4, 16 University Avenue, Macquarie University, New South Wales 2109, Australia.
Email: viji.easwar@nal.gov.au

Present address

Vijayalakshmi Easwar, National Acoustic Laboratories, Sydney, Australia.

Funding information

University of Wisconsin-Madison, Madison, WI USA 53705.

Edited by: Edmund Lalor

Abstract

Repeated stimulus presentation leads to neural adaptation and consequent amplitude reduction in vowel-evoked envelope following responses (EFRs)—a response that reflects neural activity phase-locked to envelope periodicity. EFRs are elicited by vowels presented in isolation or in the context of other phonemes such as consonants in syllables. While context phonemes could exert some forward influence on vowel-evoked EFRs, they may reduce the degree of adaptation. Here, we evaluated whether the properties of context phonemes between consecutive vowel stimuli influence adaptation. EFRs were elicited by the low-frequency first formant (resolved harmonics) and middle-to-high-frequency second and higher formants (unresolved harmonics) of a male-spoken /i/ when the presence, number and predictability of context phonemes (/s/, /a/, /j/ and /u/) between vowel repetitions varied. Monitored over four iterations of /i/, adaptation was evident only for EFRs elicited by the unresolved harmonics. EFRs elicited by the unresolved harmonics decreased in amplitude by ~16–20 nV (10%–17%) after the first presentation of /i/ and remained stable thereafter. EFR adaptation was reduced by the presence of a context phoneme, but the reduction did not change with their number or predictability. The presence of a context phoneme, however, attenuated EFRs by a degree similar to that caused by adaptation (~21–23 nV). Such a trade-off in the short- and long-term influence of context phonemes suggests that the benefit of interleaving EFR-eliciting vowels with other context phonemes depends on whether the use of consonant-vowel syllables is critical to improve the validity of EFR applications.

KEYWORDS

consonants, forward masking, frequency following response, fricatives, phase locking

Abbreviations: CV, consonant-vowel syllable; CVC, consonant-vowel-consonant syllable; EEG, electroencephalogram; EFR, envelope following response; f_0 , fundamental frequency; F1, first vowel formant; F2+, second and higher vowel formants; RM-ANOVA, repeated measures analysis of variance; SD, standard deviation.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *European Journal of Neuroscience* published by Federation of European Neuroscience Societies and John Wiley & Sons Ltd.

1 | INTRODUCTION

The vowel-evoked envelope following response (EFR) is a useful non-invasive method to assess the neural encoding of the fundamental frequency of voice (f_0). EFRs are commonly elicited by vowels that are presented in isolation or amidst other phonemes, particularly consonants, henceforth referred to as context phonemes in the present study. Although context phonemes are commonly used and are necessary in some instances, their influence on vowel-evoked EFRs is not well understood. Recent studies have focused on immediate (short-term) effects of context phonemes on EFRs (e.g. Easwar et al., 2021, 2022). In the present study, we focus on the effects of context phonemes over a longer timescale; we aimed to evaluate whether the presence and properties of context phonemes influence adaptation in vowel-evoked EFRs caused by stimulus repetition. Adaptation refers to reduced neural responsiveness that occurs with repeated presentations of the same stimulus (review by Wark et al., 2007; Pérez-González & Malmierca, 2014) and is reflected as reduced EFR amplitude over the course of the recording (Bidelman & Powers, 2018; Gorina-Careta et al., 2016).

The use of context phonemes amidst vowel stimuli has some advantages and caveats to consider. Advantages include improving test validity for hearing aid-based applications of EFRs and increasing test efficiency. Temporal characteristics of vowel stimuli embedded in consonant-vowel (CV) or consonant-vowel-consonant (CVC) syllables resemble that of running speech and are likely to facilitate accurate representation of non-linear hearing aid function (Easwar et al., 2012; Scollie & Seewald, 2002; Stelmachowicz et al., 1990; Stone & Moore, 1992). Further, if context phonemes could also elicit EFRs, that would enable gathering more data in the same recording time (e.g. Easwar, Purcell, et al., 2015a, 2015b). However, context phonemes may influence the characteristics of vowel-evoked EFRs by temporal masking, a phenomenon thought to be caused by short-term adaptation (Meddis & O'Mard, 2005). The susceptibility of vowel-evoked EFRs to temporal masking has been shown in some studies (e.g. Easwar et al., 2022; Hodge et al., 2018) but not in others (Easwar et al., 2021). EFR peaks were delayed when speech-shaped noise preceded the stimulus vowel (Hodge et al., 2018), and the amplitudes of EFRs elicited by the second and higher formants (but not the first formant) of /i/ were attenuated by 14.9 to 27.9 nV when preceded by /j/, /m/ or /i/ (Easwar et al., 2022). Together, these studies suggest that although interleaving vowel stimuli with other phonemes, particularly consonants, help improve resemblance to running speech, it may influence the interpretation and detection of EFRs when elicited by higher-frequency vowel formants.

Adaptation-related change in EFR amplitude has been quantified; however, the estimates vary and the influence of context phonemes on such changes remains unclear. EFR amplitude reduced as much as ~ 35 to 1000 nV over the course of the first 200 to 300 stimulus repetitions (Bidelman & Powers, 2018; Gorina-Careta et al., 2016). The wide range of amplitude reduction in the two EFR studies may, in part, be related to differences in the analysis. EFR amplitudes over time were either compared after averaging every 100 consecutive (non-overlapping) trials (Gorina-Careta et al., 2016) or from every trial without averaging (Bidelman & Powers, 2018). Further, differences may also be due to the context in which the vowel was presented. A CV syllable (/wa/) was used in Gorina-Careta et al. (2016), whereas the vowel stimulus was presented in isolation in Bidelman and Powers (2018). Irrespective of these methodological differences, both these studies suggest that (i) stimulus-specific adaptation, similar to that observed at cortical and subcortical levels (Anderson & Malmierca, 2012; Duque et al., 2016; Kudela et al., 2018; Malmierca et al., 2009; Ulanovsky et al., 2003; Zhao et al., 2011), is evident in EFRs too, and (ii) the degree of repetition-related attenuation of EFR amplitude could be close to or larger than the short-term masking effects caused by context phonemes preceding the EFR-eliciting vowel (discussed above). However, due to the use of *either* a vowel or a CV in these studies, it remains unclear whether context phonemes and their characteristics affect the degree of adaptation in EFRs.

The use of context phonemes between EFR-eliciting vowels and their characteristics may influence repetition-related changes in EFRs due to the auditory system's known sensitivity to stimulus history or novelty (as reflected in EFRs). Repeating the same stimulus over time as opposed to interleaving EFR stimuli with other stimuli caused a continuum of changes in listeners ranging from attenuation to enhancement of EFR amplitude at f_0 (Parbery-Clark et al., 2011; Skoe et al., 2013) or its second harmonic (Chandrasekaran et al., 2009; Slabu et al., 2012) and altering the accuracy of tracking dynamic pitch (Lau et al., 2017). Further, the amplitude of EFRs elicited by frequently occurring or equally probable stimuli tended to be enhanced compared to novel or deviant stimuli (Gnanateja et al., 2013; Slabu et al., 2012). As such, enhancements with stimulus repetition are somewhat contradictory to the adaptation-related attenuation reported in near- and far-field studies (Bidelman & Powers, 2018; Gorina-Careta et al., 2016; Prado-Gutierrez et al., 2015) and the restoration of sensitivity in adapted neurons with a change in stimulus or stimulus parameter (e.g. level or modulation frequency; Prado-Gutierrez et al., 2015). However, comparisons are challenged by differences such as in stimulus probability, averaging

method and specific conditions of enhancement observation. For example, enhancements in EFR amplitude have been evident for /da/ and /ba/ but not /wa/ (Slabu et al., 2012). Further, enhancements have been evident at the second harmonic but not at f_0 in some studies (Chandrasekaran et al., 2009; Slabu et al., 2012) and vice versa in others (Gnanateja et al., 2013). Likewise, enhancements have been demonstrated in response to the CV formant transition but not the steady-state vowel (Chandrasekaran et al., 2009; Slabu et al., 2012) and in musicians only (Parbery-Clark et al., 2011).

Two additional aspects about adaptation in EFRs remain unclear. The first aspect is the stimulus-frequency-dependent susceptibility. Frequency dependency in vowel-evoked EFRs is difficult to infer from past work due to the use of broadband vowels that provide a cumulative response at f_0 with contributions from more than one formant (Aiken & Picton, 2006; Easwar et al., 2018). Frequency specificity is an important consideration because adaptation is greater in fibres with high characteristic frequencies (>1.5 kHz) compared with those with low characteristic frequencies (in ferrets; Sumner & Palmer, 2012), and temporal masking effects are evident only for EFR stimuli above ~1200 Hz (Easwar et al., 2022). Further, the influence of context phonemes on frequency-specific repetition-related adaptation will inform EFR paradigms for clinical applications in individuals with hearing loss (e.g. Easwar, Purcell, et al., 2015b), where hearing loss degree, and therefore, the effects of hearing loss will likely vary by frequency. The second aspect is the time course of adaptation. The time course remains uncertain since prior studies either reported single-trial data that are susceptible to changes in noise or they averaged over consecutive stimuli leading to a loss of resolution in terms of stimulus repetition order/number. To seek clarity in both these aspects, in the present study, we (i) modified vowel stimuli to elicit independent EFRs from first (F1) and second and higher-frequency formants (F2+; Easwar, Purcell, et al., 2015a; Easwar et al., 2019) and (ii) used a vertical averaging approach that would maintain the vowel repetition order during averaging and reduce the impact of noise while evaluating the adaptation time course (e.g. Prado-Gutierrez et al., 2015). Vertical averaging differs from traditional averaging in situations where each trial contains several repetitions of the stimulus. Rather than averaging over all stimulus repetitions in every trial, vertical averaging averages across trials without collapsing across stimulus repetitions within a trial. Therefore, vertical averaging provides an across-trial average for each stimulus order within a trial.

In summary, we aimed to (i) evaluate the effect of presence, length and predictability of the preceding context phonemes on adaptation in vowel-evoked EFRs and (ii) evaluate frequency-specific effects of adaptation in

vowel-evoked EFRs. We hypothesised that adaptation, if present, will be the largest for EFRs elicited by vowel stimuli without any interleaving context phonemes and will be the smallest for EFRs elicited by vowel stimuli preceded by the least predictable sequence of context phonemes. Given the high-frequency bias in adaptation for tonal stimuli and greater susceptibility of F2+ EFRs than F1 EFRs to immediately preceding phonemes, we predicted larger adaptation effects for F2+ than F1 EFRs.

2 | METHOD

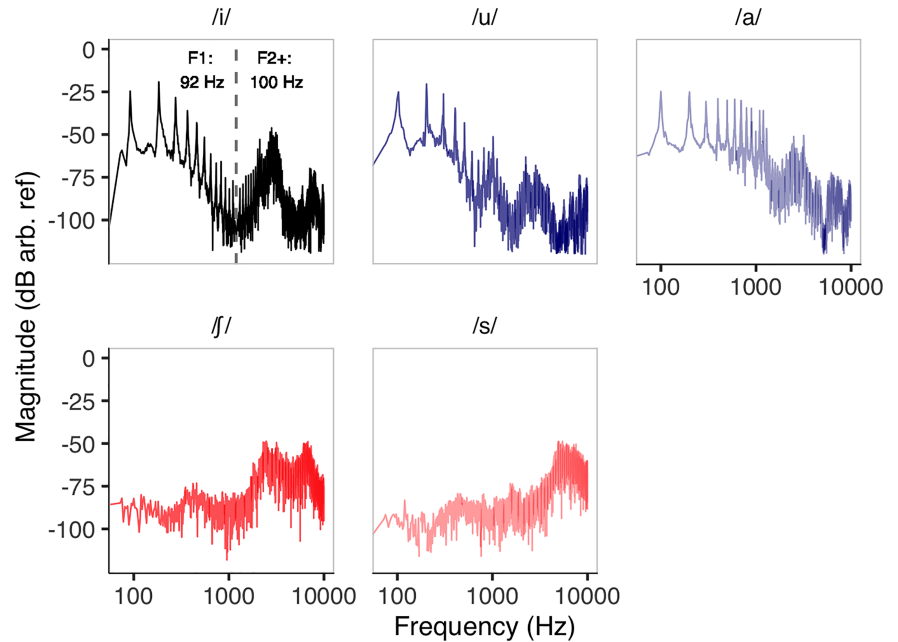
2.1 | Participants

A total of 21 young adults (mean age = 21.9 years; $SD = 2.3$; 16 females) provided written consent to participate in the study. Eligibility for participation included (i) detection of pure tones at 20 dB HL in both ears, presented using headphones (AD629, Interacoustics, Denmark), (ii) no contraindications observed in otoscopy, (iii) type A tympanogram (Titan, Interacoustics, Denmark) and (iv) no self-disclosed neurological disorders. The study protocol was approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board. Participants were either offered extra credit or compensated at \$10/h for their time.

2.2 | Stimulus

The vowel /i/ spoken by a 24-year-old male from Wisconsin, USA, was chosen to elicit EFRs. The vowel was spoken in isolation with an average f_0 of 100.51 Hz (range = 100.1–100.9 Hz). The vowel /i/ was chosen because it is commonly used in vowel-evoked EFR studies (Aiken & Picton, 2006, 2008; Choi et al., 2013; Easwar, Purcell, et al., 2015a), and the effects of preceding phoneme have been evaluated for /i/-elicited EFRs (Easwar et al., 2022). The first and second formant peak frequencies of /i/ were 228.97 and 2215.96 Hz, respectively. The phonemes /a/, /u/, /ʃ/ (“shh”) and /s/, also spoken in isolation by the same male, were chosen as the context phonemes. The five phonemes were chosen as they have been used as stimuli in prior EFR studies (Easwar, Purcell, et al., 2015a, 2015b; Easwar et al., 2020). Multiple recordings of each phoneme were made, and one of the iterations was chosen based on sound quality. Since the phonemes were produced in isolation, no co-articulation was evident in any of the phoneme productions and they were all truncated to 360 ms from their varied original length. For all phonemes, 10 ms of \sin^2 rise and fall ramps were added at the beginning and end. Spectra of phonemes are provided in Figure 1.

FIGURE 1 Spectra of the envelope following response (EFR) stimulus /i/ and other phonemes. The vertical dashed line in the first panel demarcates the two simultaneously presented formant bands F1 and F2+. The 92 and 100 Hz refer to the f_0 in F1 and F2+, respectively.



The vowel /i/ was modified to elicit two EFRs simultaneously, one from the low-frequency first formant (F1) and one from the second and higher formants (F2+). As done in previous studies for improved frequency specificity of vowel stimuli (Easwar, Purcell, et al., 2015a; Easwar et al., 2019), differentiation of the two formant bands was maintained to examine the presence and nature of adaptation in EFRs elicited by formants dominant in different spectral regions (Figure 1). The vowel was modified using the following steps to enable eliciting two EFRs simultaneously: (i) The average f_0 of the vowel was reduced by 8.57 Hz in Praat, (ii) the F1 was obtained by low-pass filtering the lowered- f_0 vowel at 1140 Hz, (iii) the F2+ was obtained by high-pass filtering the original f_0 vowel at 1250 Hz and (iv) the F1 and F2+ were combined without changes in their relative levels. The F1 consisted of the first 12 harmonics, while the F2+ consisted of the 13th and higher harmonics.

Four stimulus conditions were created. (i) No-context: The stimulus /i/ was presented without any context, (ii) single-context: The stimulus /i/ was preceded by the context phoneme /s/ (i.e. /si/), (iii) multiple-context: The stimulus /i/ was preceded by the same sequence of context phonemes /uʃas/ (i.e. /uʃasi/), and (iv) random-context: The stimulus /i/ was preceded by one of the four sequences created with the same four context phonemes (i.e. /aʃusi/, /ʃausi/, /ʃuasi/ and /uaʃsi/). The random nature of context in the “random-context” condition refers to the presentation of the three phonemes (/ʃ/, /u/ and /a/). Since the immediately preceding context phoneme could influence the amplitude of EFRs, especially the ones elicited by F2+ (Easwar et al., 2022), the same context phoneme /s/ was used in all conditions with a

context phoneme. Maintaining the same context phoneme improved the separation of context influence in the short-term and the long-term, the latter being the goal of the present study. All the vowels were equated in root-mean-square (RMS) level in Praat. The RMS level of the fricatives was 10 dB lower than that of the vowels. The difference in level between the vowels and the fricatives approximated relative levels in naturally produced CV syllables (Easwar, Purcell, et al., 2015a; Easwar et al., 2020). Stimulus waveforms are shown in Figure 2.

The first half of each stimulus trial consisted of the stimulus (sequence in the case of single-, multiple- and random-contexts) presented four times without interstimulus intervals. The second half of each stimulus trial consisted of the same four stimulus iterations in the opposite polarity. A silent interval of 500 ms was included at the beginning of each half trial in all conditions. A total of 250 stimulus trials were presented in all conditions. The four iterations of the /i/ within each trial will henceforth be referred as per the order in which they were presented (i.e. i1, i2, i3 and i4). The duration of recording varied by condition; recording lasted for ~16 min for the no-context condition, ~28 min for the single-context condition and ~64 min for the multiple- and random-context conditions.

2.3 | Stimulus presentation and EFR recording

Stimulus presentation and EFR recording was controlled by the Auditory Research Lab Audio software (Goodman, 2017; github.com/myKungFu/ARLas) developed at the

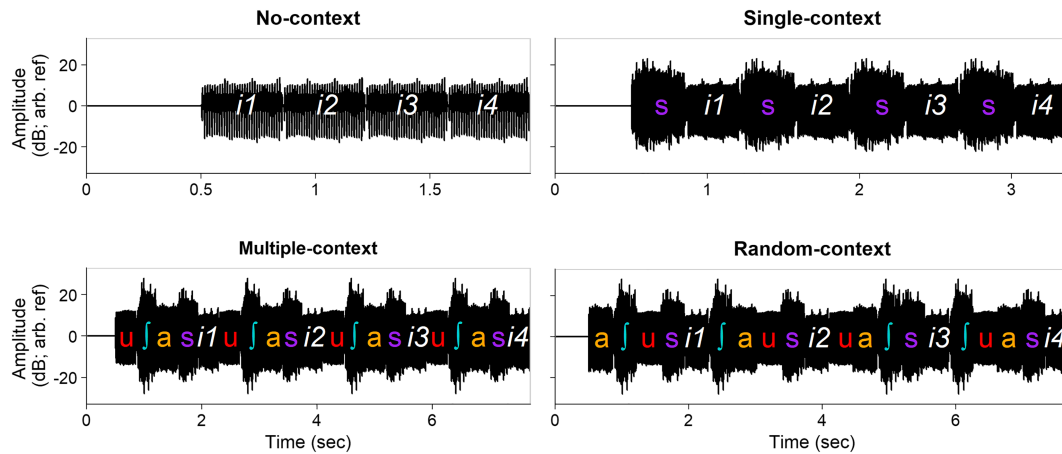


FIGURE 2 Waveform of each stimulus

University of Iowa. Digital-to-analogue conversion of the stimulus was completed by Fireface UCFx+ (RME, Haimhausen, Germany). The sampling rate of both the stimulus and electroencephalogram (EEG) was 96,000 samples/s. EEG was later downsampled to 8000 samples/s. The stimulus was presented in a randomly chosen test ear using an ER2 insert earphone (Etymotic Research, IL) at an RMS level of 70 dB SPL measured over the entire duration of /i/. Stimulus level calibration was completed in an ear simulator (Brüel & Kjør ear simulator Type 4517) using a sound conditioner (Type 1704; Nærum, Denmark). The test ear was the right ear for 10 participants.

EEG was recorded using one of the channels of the Intelligent Hearing Systems OptiAmp with sintered Ag-AgCl electrodes placed at the vertex (Cz; non-inverting), the nape (inverting) and the collar bone (ground). Electrode sites were prepped with alcohol wipe and NuPrep to achieve impedances of <3 k Ω . Analogue-to-digital conversion of the EEG was completed by the RME Fireface UCFx+. During EEG recordings, participants were seated in a comfortable chair that reclined and watched a silent movie of choice with subtitles. Recordings were completed in an electromagnetically shielded double-walled sound booth.

2.4 | EFR analysis

Analysis steps were similar to previous EFR studies using naturally spoken vowels as the stimulus (Choi et al., 2013; Easwar, Beamish, et al., 2015; Easwar, Purcell, et al., 2015a). The analysis window of /i/ entailed the central 350 ms and excluded 5 ms on both ends. EFRs were estimated with a Fourier analyser that utilized the original and lowered f_0 time courses obtained using Praat (Choi et al., 2013; Easwar, Beamish, et al., 2015; Easwar,

Purcell, et al., 2015a; Easwar et al., 2021). Reference cosine and sine sinusoids were created using the f_0 time courses. EEG was multiplied with the reference sinusoids, after accounting for a 10-ms brainstem processing delay (Choi et al., 2013; Easwar et al., in press), to obtain real and imaginary components of the EFR. Each component was averaged over the entire 350-ms-long window to estimate the amplitude and phase of EFRs. Residual noise was calculated from EEG amplitudes at 14 frequencies surrounding the original and lowered f_0 , except for 120 Hz. EFR amplitudes were unbiased to reduce the influence of noise (Picton et al., 2005). When the ratio of EFR to noise amplitude exceeded 1.25, the EFR amplitude was divided by an overestimation factor, computed as $1 + 0.965 * e^{-1.348 * X} + 0.078 * e^{-0.285 * X}$, where X refers the ratio between EFR and noise amplitude (Picton et al., 2005).

2.5 | Data exclusion

Participants were excluded based on residual noise levels. Two participants were excluded as their residual noise was greater than the third quartile+1.5*interquartile range of the noise computed for the group, in at least three of 32 (10%) recordings. Data from the remaining 19 participants were included in the statistical analysis.

2.6 | Statistical analysis

A three-way repeated measure analysis of variance (RM-ANOVA) was completed with condition (no-context, single-context, multiple-context and random-context), order (i1, i2, i3 and i4) and formant (F1, F2+) as the three within-subject factors. Following significant effects in the main model, pairwise comparisons using paired

t tests were completed. To account for inflation of alpha error due to multiple tests, *p*-values were corrected using the false-discovery rate approach (Benjamini & Hochberg, 1995). Corrected *p*-values are reported throughout the manuscript, and therefore, $p < 0.05$ are to be interpreted as statistically significant. All statistical analyses were completed in R (v 4.0.4; R Core Team, 2021).

2.7 | Results

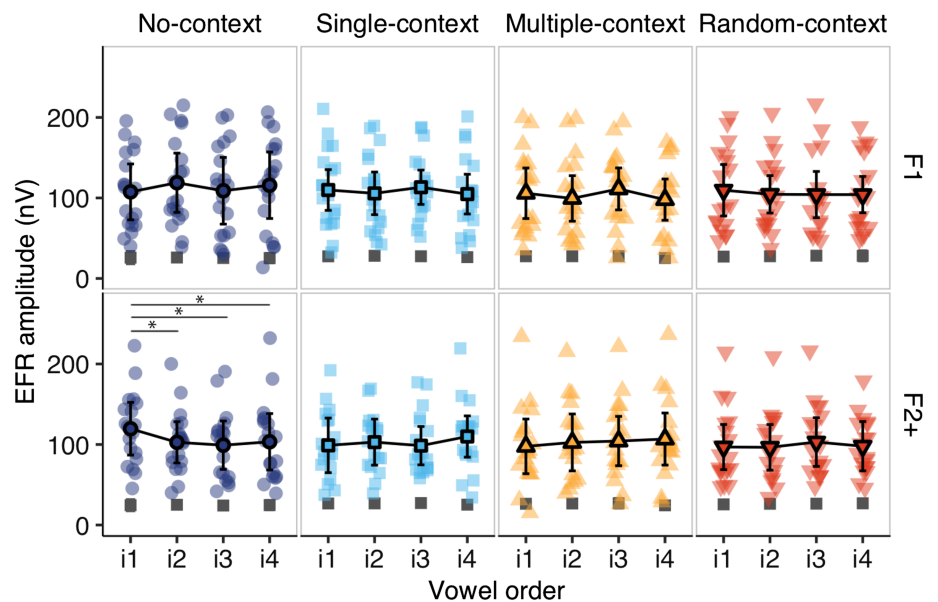
Figure 3 illustrates EFR amplitudes as a function of vowel order for F1 and F2+ stimuli within each condition. The RM-ANOVA indicated a significant three-way interaction between condition, vowel order and formant ($F[9,162] = 2.38$, $p = 0.015$, $\text{partial-}\eta^2 = 0.14$) suggesting that the EFR amplitude varied as a function of all three factors. None of the main effects were significant in the three-way RM-ANOVA. That is, EFR amplitudes did not vary by the context condition (averaged over all vowel orders; $F[3,54] = 1.63$, $p = 0.194$, $\text{partial-}\eta^2 = 0.08$), the formant eliciting the EFR ($F[1,18] = 0.26$, $p = 0.616$, $\text{partial-}\eta^2 = 0.01$) as well as the vowel order (averaged across all context conditions; $F[3,54] = 0.21$, $p = 0.892$, $\text{partial-}\eta^2 = 0.01$). Further, none of the two-way interactions were significant (condition \times vowel: $F[9,162] = 0.99$, $p = 0.450$, $\text{partial-}\eta^2 = 0.05$; condition \times formant: $F[3,54] = 0.52$, $p = 0.668$, $\text{partial-}\eta^2 = 0.03$; vowel \times formant; $F[3,54] = 0.56$, $p = 0.646$, $\text{partial-}\eta^2 = 0.03$).

Posthoc analyses following the significant three-way interaction were completed to assess the presence of adaptation as well as the effect of varying preceding

context. To evaluate the presence of adaptation, paired *t* tests were completed to compare EFR amplitude between multiple vowel orders within each stimulus. For F1 EFRs, none of the pairwise comparisons were statistically significant suggesting no reduction in EFR amplitude with stimulus repetition (all $p > 0.05$). For F2+ EFRs, pairwise comparisons were statistically significant for the no-context condition only. In the no-context condition, the amplitude of EFRs elicited by i1 was significantly larger than those elicited by i2, i3 and i4, by mean differences of 16.9 ($SD = 25.9$), 20.4 ($SD = 20.4$) and 16.2 nV ($SD = 26.7$), respectively (paired *t* test between i1 and i2: $t[18] = 2.84$, $p = 0.032$, Cohen's $d = 0.65$; between i1 and i3: $t[18] = 4.36$, $p = 0.002$, Cohen's $d = 1.0$; between i1 and i4: $t[18] = 2.65$, $p = 0.032$, Cohen's $d = 0.61$). EFR amplitude did not vary between i2, i3 and i4 in the no-context condition (all $p > 0.05$). Likewise, EFR amplitudes did not vary as a function of vowel order in single-, multiple- and random-context conditions (all $p > 0.05$). In terms of percent reduction relative to the amplitude for i1 in the no-context condition, F2+ EFRs reduced by 10.5% ($SD = 24.6$), 16.8% ($SD = 15.5$) and 13.2% ($SD = 23.6$) by the second, third and fourth repetition, on average. In summary, a reduction in EFR amplitude with stimulus repetition was evident only for the EFR elicited by F2+ and only in conditions without a preceding context phoneme.

Figure 4 illustrates EFR amplitudes as a function of context condition at each vowel order for both F1 and F2+ vowel formants. To assess the effect of the different types of preceding context on EFRs, paired *t* tests were completed on EFR amplitudes between conditions at each vowel order. EFRs elicited by F1 did not vary between conditions at any of the vowel positions

FIGURE 3 Envelope following response (EFR) amplitude as a function of vowel order in each condition for F1 and F2+ stimuli. Coloured symbols with black outline represent group means. Coloured symbols represent individual data. Filled grey squares represent group mean noise amplitude. Error bars represent within-subject standard deviation. * indicates a statistically significant pairwise comparison.



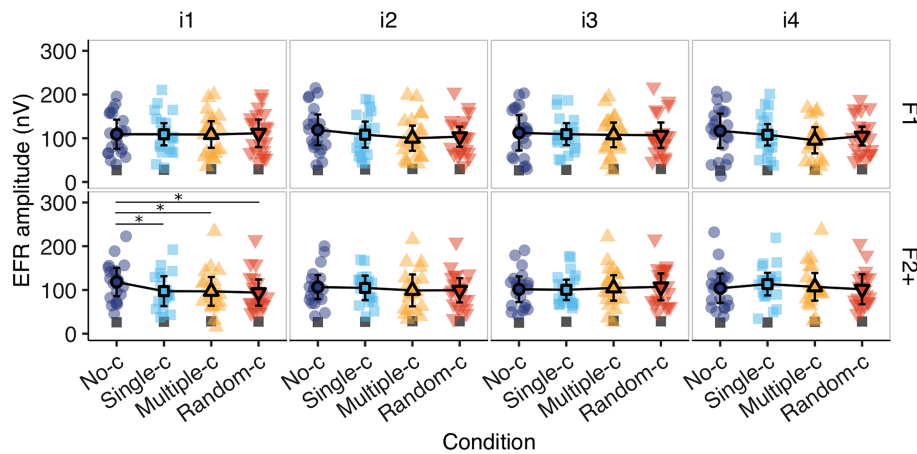


FIGURE 4 Envelope following response (EFR) amplitude as a function of condition at each vowel order for F1 and F2+ stimuli. “c” in the x-axis refers to condition. Coloured symbols with black outline represent group means. Coloured symbols represent individual data. Filled grey squares represent group mean noise amplitude. Error bars represent within-subject standard deviation. * indicates a statistically significant pairwise comparison.

(all $p > 0.05$). In contrast, the amplitude of EFRs elicited by F2+ varied as a function of condition. However, this was only evident for the first vowel i1. At i1, the EFR amplitude was significantly larger in the no-context condition compared with single-, multiple- and random-context conditions by mean differences of 20.6 ($SD = 26.7$), 21.9 ($SD = 34.1$) and 22.7 ($SD = 35.4$) nV, respectively (paired t test between no-context and single-context: $t[18] = 3.38$, $p = 0.020$, Cohen’s $d = 0.77$; no-context and multiple-context: $t[18] = 2.80$, $p = 0.024$, Cohen’s $d = 0.64$; no-context and random-context: $t[18] = 2.79$, $p = 0.024$, Cohen’s $d = 0.64$). EFR amplitudes did not vary between the different context conditions at any of the other vowels (all $p > 0.05$). In terms of percent reduction relative to the amplitude for i1 in the no-context condition, F2+ EFRs reduced by 15.6% ($SD = 24.1$), 16.3% ($SD = 34.4$) and 15.2% ($SD = 29.6$) in the single-context, multiple-context and random-context, on average. In summary, the presence of a preceding context led to a reduction in the amplitude of EFRs elicited by /i/ F2+; however, this was evident only for the first stimulus presentation.

Changes in EFR amplitude with vowel repetition or preceding context summarised above could not be explained by changes or differences in noise amplitude. Similar to EFR amplitude, a three-way RM-ANOVA was completed for the noise amplitude. ANOVA results suggested that the noise amplitudes varied significantly only by the vowel formant. Residual noise in F1 EFRs was larger than that in F2+ EFRs by 1.1 nV ($SD = 0.99$), on average. No other main or interaction effects were statistically significant (all $p > 0.05$).

2.8 | Discussion

The purpose of the present study was to evaluate if the presence, length and predictability of context phonemes

preceding the EFR-eliciting vowel influenced adaptation in EFRs. We found that interleaving vowel stimuli with other phonemes reduces adaptation-related EFR attenuation, irrespective of the length and predictability of the other phonemes.

2.9 | Reduction in response amplitude with stimulus repetition was evident only for EFRs elicited by the second and higher vowel formants (unresolved harmonics)

A novel aspect of this study is the frequency-specific investigation of adaptation in EFRs elicited by vowel stimuli. Our results indicate a reduction in response amplitude with stimulus repetition only for EFRs elicited by F2+ (Figure 3). Such frequency dependency is consistent with greater adaptation-related attenuation evident at high characteristic frequencies (>1.5 kHz) in ferrets (Sumner & Palmer, 2012). A speculated cause for frequency dependency is the presence of phase-locking at the lower frequencies. With dual- f_0 vowels used in the present study, the EFR elicited by /i/ F1 was predominantly initiated by harmonics resolved in the cochlea, whereas the EFR elicited by /i/ F2+ was predominantly initiated by unresolved harmonics (Micheyl & Oxenham, 2004; Laroche et al., 2013; Easwar, Beamish, et al., 2015; Easwar et al., 2019). Phase-locked responses to the stimulus fine structure or vowel harmonics have been detected at frequencies as high as 1120 (Bidelman & Powers, 2018) to 1500 Hz (Aiken & Picton, 2008). These upper limits would suggest that the lower-frequency harmonics in the F1 stimulus of the present study likely facilitated phase-locking to the individual harmonics and that may have played a role in the lack of adaptation or change observed with stimulus repetition for F1 EFRs.

2.10 | EFRs elicited by the second and higher formants of /i/ are influenced only by the immediately preceding fricative

F2+ EFRs were attenuated by an average of ~ 21 to 23 nV when the stimulus /i/ was preceded by /s/ (Figure 4). The influence of immediately preceding phoneme on F2+ EFRs is consistent with our previous work where a preceding /j/, presented 15 dB lower than EFR-eliciting vowel /i/, reduced response amplitude by 16.04 nV ($SD = 21.42$), on average (Easwar et al., 2022). F2+ EFRs were more susceptible than F1 EFRs to the influence of preceding phonemes possibly because (i) /i/ F2+ was ~ 18 dB lower than /i/ F1 and (ii) the spectral overlap or similarity was greater between /s/ and /i/ F2+ than between /s/ and F1 due to the high-frequency emphasis of fricatives. As shown in several studies investigating temporal masking, larger level differences and greater spectral similarity increase the probability and extent of temporal (forward) masking (Abbas & Gorga, 1981; Gao & Berrebi, 2015; Kramer & Teas, 1982; Lasky & Rupert, 1982; Nelson et al., 2009). Of note is the lack of any masking effects of surrounding phonemes on /i/-elicited EFRs in our other recent study (Easwar et al., 2021); lack of differentiation of F1 and F2+ EFRs and/or shorter duration of interleaving phonemes in the previous study may have contributed to the differences.

Our data also suggest that EFRs were likely insensitive to stimulus history earlier than the immediately preceding phoneme. That is, varied length or predictability of the sequence of phonemes (at constant probability of occurrence) did not influence vowel-evoked EFRs over time. This is supported by two comparisons: first, the lack of differences in EFR amplitude between the single-context, multiple-context and random-context conditions at i1 (Figure 4) and second, the similar degree of differences between the no-context condition and each of the three conditions at i1 (Figure 4). These results are somewhat inconsistent with Skoe et al. (2013), where EFRs elicited by tones were found to be larger in amplitude when stimuli were presented in randomised stimulus sequences compared to when tone pairs were fixed in sequence and repeated. Differences could be largely methodological and possibly participant-related. First, the timing of the “i” was predictable in the present study. That is, “i” always occurred after four other phonemes to control for the difference in temporal gap between two iterations and was always preceded by “s” to control for short-term temporal masking effects, whereas the timing of tones in Skoe et al. (2013) was less or not predictable. The second reason may relate to the degree of randomness introduced between the conditions being compared. In the present study, the multiple-context repeated the same exact sequence, while Skoe et al. presented stimulus

(tone) doublets amidst a random sequence of other tones. The third reason may be individual variability related to musical training (Parbery-Clark et al., 2011) or ability to detect pattern in stimuli (Skoe et al., 2013), both of which were not probed in the present study.

2.11 | Interleaving EFR vowel stimuli with other phonemes reduced the degree of adaptation

A reduction in F2+ EFRs with vowel repetition was evident only when there were no phonemes between the vowel repetitions (no-context condition; Figure 3). The reduction in amplitude with repeated exposure to the same stimulus is consistent with adaptation previously observed with vowel-evoked EFRs at f_0 (Gorina-Careta et al., 2016; Bidelman & Powers, 2018). The reduction in EFR amplitude of 16–20 nV (about 10%–17%) mainly occurred after the first vowel repetition (i.e. EFR amplitude was similar between the second, third and fourth repetitions). Although the rapid reduction in EFR amplitude during the initial repetitions is consistent with previous studies (Gorina-Careta et al., 2016; Bidelman & Powers, 2018), the number of repetitions over which the rapid reduction occurred differ across studies. Single-trial data from Bidelman and Powers (2018) demonstrate a steep reduction of ~ 1 μ V (1000 nV; about 30% drop) in the first 200 of 2000 trials—an estimate much larger than the present study as well as Gorina-Careta et al. (2016), who reported an average reduction of 35 nV (about 18% drop) over the first 300 trials when EFR amplitude was estimated from non-overlapping 100-trial averages (i.e. 1–100, 101–200 etc.). As acknowledged by Bidelman and Powers (2018), the influence of change in residual noise over time on the large adaptation-related attenuation cannot be ruled out in their study. Although the degree of adaptation is more comparable between the present study and Gorina-Careta et al. (2016), the time, or the number of repetitions by which stable amplitude is achieved, is difficult to compare due to the different types of averaging. While the vertical averaging in the present study provides better stimulus-order resolution in the time course of adaptation, it does not evaluate past four repetitions. By contrast, data by Gorina-Careta et al. (2016) allowed for monitoring the time course over a longer period albeit with lower stimulus-order resolution. Based on data from Gorina-Careta et al. (2016), it is possible that further changes in amplitude will be evident past four repetitions.

The present study data suggest that interleaving vowel stimuli with context phonemes presents a trade-off between attenuation due to temporal masking and reduction or avoidance of adaptation. While the presence of

context phonemes caused ~21–23 nV of attenuation in EFR amplitude on average, it avoided ~16–20 nV of adaptation-related attenuation. Given the similarity in magnitude between the two effects and the additional benefit of resembling temporal characteristics of running speech with the use of CV syllables, we infer that interleaving EFR stimuli with other phonemes is favourable, especially for aided (i.e. with hearing aid) applications in individuals with hearing loss. The favourable situation exists even for one phoneme interleaving the EFR-eliciting vowel stimuli in a predictable manner.

The present study did not incorporate interstimulus intervals between repeated stimuli in an attempt to simulate stimulus designs in previous studies that have implications for aided applications (e.g. Easwar, Purcell, et al., 2015b). Given the similarity in the degree of adaptation between the present study and the study by Gorina-Careta et al. (2016), who used interstimulus intervals of about 196 ms, we speculate that the lack of interstimulus interval was not a confound in our measurements.

Introducing a context phoneme between two repetitions of the EFR-eliciting /i/ necessarily increased the duration between two consecutive repetitions of /i/. Therefore, the effective inter-repetition duration increased between the no-context condition, the single-context condition and the conditions that included multiple contexts presented in a predictable or randomised order. It is possible that the change in the inter-repetition duration contributed to our inference that the use of a context phoneme reduces adaptation in EFRs. However, our data in Figure 3 suggest that the inter-repetition duration was not the main or sole factor influencing adaptation in EFRs; the degree of adaptation was no greater in conditions with a single-context compared with that in multiple- or random-context conditions with five times the inter-repetition duration as in the single-context condition. We thus infer that the reduction in adaptation is predominantly due to the presence of at least one context phoneme.

3 | SUMMARY AND CONCLUSIONS

The present study findings suggest that EFRs elicited by low-frequency first vowel formants are neither susceptible to repeated stimulation and subsequent adaptation nor are they susceptible to the nature of preceding stimuli or context phonemes. In comparison, EFRs elicited by the second and higher vowel formants are not only susceptible to attenuation from repeated stimulus presentation but also sensitive to the immediately preceding phoneme. Although the immediately preceding phoneme attenuates EFRs elicited by the higher-frequency formants, the presence of even one such phoneme between

repetitions of /i/ reduces the degree of adaptation. Using a single phoneme between EFR-eliciting vowels provided no lesser benefit than using multiple phonemes in a predictable or pseudorandom sequence. Given such a trade-off in the influence of context phonemes, we conclude that the benefit of interleaving EFR-eliciting vowels with other phonemes likely depends on whether the use of CV syllables to simulate running speech is desired for the intended application of EFRs.

ACKNOWLEDGEMENTS

The study was funded by the New Investigator Research Grant by American Speech-Language-Hearing Foundation awarded to VE. The authors thank Madeline Salutz for assistance with data collection, Veronika Mäk for assistance with preparing figures and proof reading, and David Purcell for sharing basic analysis scripts.

CONFLICT OF INTEREST

none to declare.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ejn.15768>.

DATA AVAILABILITY STATEMENT

Data will be made available in response to email requests made to the corresponding author.

ORCID

Vijayalakshmi Easwar  <https://orcid.org/0000-0001-7944-3012>

REFERENCES

- Abbas, P. J., & Gorga, M. P. (1981). AP responses in forward-masking paradigms and their relationship to responses of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, 69, 492–499. <https://doi.org/10.1121/1.385477>
- Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology and Neurotology*, 11, 213–232. <https://doi.org/10.1159/000092589>
- Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, 245, 35–47. <https://doi.org/10.1016/j.heares.2008.08.004>
- Anderson, L. A., & Malmierca, M. S. (2012). The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. *European Journal of Neuroscience*, 37, 52–62. <https://doi.org/10.1111/ejn.12018>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bidelman, G., & Powers, L. (2018). Response properties of the human frequency-following response (FFR) to speech and

- non- speech sounds: Level dependence, adaptation and phase-locking limits. *International Journal of Audiology*, 57, 665–672. <https://doi.org/10.1080/14992027.2018.1470338>
- Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T. G., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, 64, 311–319. <https://doi.org/10.1016/j.neuron.2009.10.006>
- Choi, J. M., Purcell, D. W., Coyne, J.-A. M., & Aiken, S. J. (2013). Envelope following responses elicited by English sentences. *Ear and Hearing*, 34, 637–650. <https://doi.org/10.1097/AUD.0b013e31828e4dad>
- Duque, D., Wang, X., Nieto-Diego, J., Krumbholz, K., & Malmierca, M. S. (2016). Neurons in the inferior colliculus of the rat show stimulus-specific adaptation for frequency, but not for intensity. *Nature Neuroscience*, 6, 1–15.
- Easwar, V., Aiken, S., Beh, K., McGrath, E., Galloy, E., Scollie, S., & Purcell, D. (in press). Variability in the estimated amplitude of vowel-evoked envelope following responses caused by assumed neurophysiologic processing delays. *Journal of Association for Research in Otolaryngology*. <https://doi.org/10.1007/s10162-022-00855-1>
- Easwar, V., Banyard, A., Aiken, S. J., & Purcell, D. W. (2018). Phase-locked responses to the vowel envelope vary in scalp-recorded amplitude due to across-frequency response interactions. *European Journal of Neuroscience*, 48, 3126–3145. <https://doi.org/10.1111/ejn.14161>
- Easwar, V., Beamish, L., Aiken, S. J., Choi, J. M., Scollie, S. D., & Purcell, D. W. (2015). Sensitivity of envelope following responses to vowel polarity. *Hearing Research*, 320, 38–50. <https://doi.org/10.1016/j.heares.2014.11.008>
- Easwar, V., Birstler, J., Harrison, A., Scollie, S. D., & Purcell, D. W. (2020). The accuracy of envelope following responses in predicting speech audibility. *Ear and Hearing*, 41, 1732–1746. <https://doi.org/10.1097/AUD.0000000000000892>
- Easwar, V., Boothalingam, S., & Wilson, E. (2022). Sensitivity of vowel-evoked envelope following responses to spectra and level of preceding phoneme context. *Ear and Hearing*, 43, 1327–1335. <https://doi.org/10.1097/AUD.0000000000001190>
- Easwar, V., Bridgwater, E., & Purcell, D. (2021). The influence of vowel identity, vowel production variability, and consonant environment on envelope following responses. *Ear and Hearing*, 42, 662–672. <https://doi.org/10.1097/AUD.0000000000000966>
- Easwar, V., Purcell, D. W., Aiken, S. J., Parsa, V., & Scollie, S. D. (2015a). Effect of stimulus level and bandwidth on speech-evoked envelope following responses in adults with Normal hearing. *Ear and Hearing*, 36, 619–634. <https://doi.org/10.1097/AUD.0000000000000188>
- Easwar, V., Purcell, D. W., Aiken, S. J., Parsa, V., & Scollie, S. D. (2015b). Evaluation of speech-evoked envelope following responses as an objective aided outcome measure: Effect of stimulus level, bandwidth, and amplification in adults with hearing loss. *Ear and Hearing*, 36, 635–652. <https://doi.org/10.1097/AUD.0000000000000199>
- Easwar, V., Purcell, D. W., & Scollie, S. D. (2012). Electroacoustic comparison of hearing aid output of phonemes in running speech versus isolation: Implications for aided cortical auditory evoked potentials testing. *International Journal of Otolaryngology*, 2012, 1–10, 518202. <https://doi.org/10.1155/2012/518202>
- Easwar, V., Scollie, S. D., & Purcell, D. W. (2019). Investigating potential interactions between envelope following responses elicited simultaneously by different vowel formants. *Hearing Research*, 380, 35–45. <https://doi.org/10.1016/j.heares.2019.05.005>
- Gao, F., & Berrebi, A. S. (2015). Forward masking in the medial nucleus of the trapezoid body of the rat. *Brain Structure and Function*, 221, 1–15.
- Gnanateja, G. N., Ranjan, R., Firdose, H., Sinha, S. K., & Maruthy, S. (2013). Acoustic basis of context dependent brainstem encoding of speech. *Hearing Research*, 304, 28–32. <https://doi.org/10.1016/j.heares.2013.06.002>
- Gorina-Careta, N., Zarnowiec, K., Costa-Faidella, J., & Escera, C. (2016). Timing predictability enhances regularity encoding in the human subcortical auditory pathway. *Scientific Reports*, 6, 37405. <https://doi.org/10.1038/srep37405>
- Hodge, S. E., Menezes, D. C., Brown, K. D., & Grose, J. H. (2018). Forward masking of the speech-evoked auditory brainstem response. *Otology & Neurotology*, 39, 150–157. <https://doi.org/10.1097/MAO.0000000000001657>
- Kramer, S. J., & Teas, D. C. (1982). Forward masking of auditory nerve (N1) and brainstem (wave V) responses in humans. *The Journal of the Acoustical Society of America*, 72, 795–803. <https://doi.org/10.1121/1.388186>
- Kudela, P., Boatman-Reich, D., Beeman, D., & Anderson, W. S. (2018). Modeling neural adaptation in auditory cortex. *Front Neural Circuit*, 12, 72. <https://doi.org/10.3389/fncir.2018.00072>
- Laroche, M., Dajani, H. R. R., Prevost, F., & Marcoux, A. M. (2013). Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. *Ear and Hearing*, 34, 63–74. <https://doi.org/10.1097/AUD.0b013e31826119a1>
- Lasky, R. E., & Rupert, A. L. (1982). Temporal masking of auditory evoked brainstem responses in human newborns and adults. *Hearing Research*, 6, 315–334. [https://doi.org/10.1016/0378-5955\(82\)90063-6](https://doi.org/10.1016/0378-5955(82)90063-6)
- Lau, J. C. Y., Wong, P. C. M., & Chandrasekaran, B. (2017). Context-dependent plasticity in the subcortical encoding of linguistic pitch patterns. *Journal of Neurophysiology*, 117, 594–603. <https://doi.org/10.1152/jn.00656.2016>
- Malmierca, M. S., Cristaudo, S., Perez-Gonzalez, D., & Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *The Journal of Neuroscience*, 29, 5483–5493. <https://doi.org/10.1523/JNEUROSCI.4153-08.2009>
- Meddis, R., & O'Mard, L. P. (2005). A computer model of the auditory-nerve response to forward-masking stimuli. *The Journal of the Acoustical Society of America*, 117, 3787–3798. <https://doi.org/10.1121/1.1893426>
- Micheyl, C., & Oxenham, A. J. (2004). Sequential F0 comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms. *The Journal of the Acoustical Society of America*, 116, 3038–3050. <https://doi.org/10.1121/1.1806825>
- Nelson, P. C., Smith, Z. M., & Young, E. D. (2009). Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual

- masking. *The Journal of Neuroscience*, 29, 2553–2562. <https://doi.org/10.1523/JNEUROSCI.5359-08.2009>
- Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia*, 49, 3338–3345. <https://doi.org/10.1016/j.neuropsychologia.2011.08.007>
- Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: An overview. *Frontiers in Integrative Neuroscience*, 8, number 19. <https://doi.org/10.3389/fnint.2014.00019>
- Picton, T. W., Dimitrijevic, A., Perez-Abalo, M.-C., & Roon, P. V. (2005). Estimating audiometric thresholds using auditory steady-state responses. *Journal of the American Academy of Audiology*, 16, 140–156. <https://doi.org/10.3766/jaaa.16.3.3>
- Prado-Gutierrez, P., Castro-Fariñas, A., Morgado-Rodriguez, L., Velarde-Reyes, E., Martínez, A. D., & Martínez-Montes, E. (2015). Habituation of auditory steady state responses evoked by amplitude-modulated acoustic signals in rats. *Audiology Research*, 5, 21–29. <https://doi.org/10.4081/audiores.2015.113>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for statistical computing. <https://www.R-project.org/>
- Scollie, S. D., & Seewald, R. C. (2002). Evaluation of electroacoustic test signals I: Comparison with amplified speech. *Ear and Hearing*, 23, 477–487. <https://doi.org/10.1097/00003446-200210000-00009>
- Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience*, 243, 104–114. <https://doi.org/10.1016/j.neuroscience.2013.03.009>
- Slabu, L., Grimm, S., & Escera, C. (2012). Novelty detection in the human auditory brainstem. *Journal of Neuroscience Methods*, 32, 1447–1452. <https://doi.org/10.1523/JNEUROSCI.2557-11.2012>
- Stelmachowicz, P. G., Lewis, D. E., Seewald, R. C., & Hawkins, D. B. (1990). Complex and pure-tone signals in the evaluation of hearing-aid characteristics. *Journal of Speech and Hearing Research*, 33, 380–385. <https://doi.org/10.1044/jshr.3302.380>
- Stone, M. A., & Moore, B. C. J. (1992). Syllabic compression - effective compression ratios for signals modulated at different rates. *British Journal of Audiology*, 26, 351–361. <https://doi.org/10.3109/03005369209076659>
- Sumner, C. J., & Palmer, A. R. (2012). Auditory nerve fibre responses in the ferret. *European Journal of Neuroscience*, 36, 2428–2439. <https://doi.org/10.1111/j.1460-9568.2012.08151.x>
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6, 391–398. <https://doi.org/10.1038/nn1032>
- Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. *Current Opinion in Neurobiology*, 17, 423–429. <https://doi.org/10.1016/j.conb.2007.07.001>
- Zhao, L., Liu, Y., Shen, L., Feng, L., & Hong, B. (2011). Stimulus-specific adaptation and its dynamics in the inferior colliculus of rat. *Neuroscience*, 181, 163–174. <https://doi.org/10.1016/j.neuroscience.2011.01.060>

How to cite this article: Easwar, V., & Chung, L. (2022). The influence of phoneme contexts on adaptation in vowel-evoked envelope following responses. *European Journal of Neuroscience*, 56(5), 4572–4582. <https://doi.org/10.1111/ejn.15768>