NEUROSCIENCE
RESEARCH ARTICLE
Z. Xie et al. / Neuroscience 384 (2018) 64–75

Taking Attention Away from the Auditory Modality: Context-dependent Effects on Early Sensory Encoding of Speech

Zilong Xie, [a] **Rachel Reetzke** [a] **and Bharath Chandrasekaran** [a,b,c,d,e]*

[a] *Department of Communication Sciences and Disorders, The University of Texas at Austin, Austin, TX 78712, USA*

[b] *Department of Psychology, The University of Texas at Austin, Austin, TX 78712, USA*

[c] *Department of Linguistics, The University of Texas at Austin, Austin, TX 78712, USA*

[d] *Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA*

[e] *Institute for Mental Health Research, The University of Texas at Austin, Austin, TX 78712, USA*

Abstract—Increasing visual perceptual load can reduce pre-attentive auditory cortical activity to sounds, a reflection of the limited and shared attentional resources for sensory processing across modalities. Here, we demonstrate that modulating visual perceptual load can impact the early sensory encoding of speech sounds, and that the impact of visual load is highly dependent on the predictability of the incoming speech stream. Participants (n = 20, 9 females) performed a visual search task of high (target similar to distractors) and low (target dissimilar to distractors) perceptual load, while early auditory electrophysiological responses were recorded to native speech sounds. Speech sounds were presented either in a 'repetitive context', or a less predictable 'variable context'. Independent of auditory stimulus context, pre-attentive auditory cortical activity was reduced during high visual load, relative to low visual load. We applied a data-driven machine learning approach to decode speech sounds from the early auditory electrophysiological responses. Decoding performance was found to be *poorer* under conditions of high (relative to low) visual load, when the incoming acoustic stream was predictable. When the auditory stimulus context was less predictable, decoding performance was substantially *greater* for the high (relative to low) visual load conditions. Our results provide support for shared attentional resources between visual and auditory modalities that substantially influence the early sensory encoding of speech signals in a context-dependent manner.

Key words: visual perceptual load, auditory predictability, frequency-following response, machine learning.

INTRODUCTION

At any waking moment, our brains are bombarded with sensory information from multiple modalities. The central nervous system is constantly extracting meaningful patterns or regularities from the incoming stimuli (e.g., Winkler et al., 2009; Stefanics et al., 2014). Sensory systems adjust their response properties based on the stimulation context of the sensory inputs (Nelken and Ulanovsky, 2007). In addition, we often find ourselves focusing attention on a task in one modality while ignoring information from other sensory modalities. This raises two important questions: (1) To what extent does allocating attentional resources to one modality preclude the processing of task-irrelevant stimuli in another modality? (2) To what extent does such attentional modulation interact with the processing of regularities of task-irrelevant stimuli? In the current study, we addressed these questions by examining the influence of attending to a visual task on the processing of task-unrelated auditory stimuli with varying predictability.

Per *load theory*, visual and auditory processing share capacity-limited neural resources, and the depletion of resources by one modality diminishes available resources to the other modality (Lavie, 2005; Klemen et al., 2009). Behavioral studies have provided robust support for the load theory, revealing the impact of modulating visual load on audition. For example, crossmodal studies show that visual tasks of high perceptual load are associated with a reduced detection sensitivity of task-irrelevant auditory stimuli, demonstrating load-induced *inattentional deafness* (Macdonald and Lavie, 2011; Raveh and Lavie, 2015). The neural mechanisms underlying these modulatory influences are still under

*Correspondence to: B. Chandrasekaran, The University of Texas at Austin, 2504A Whitis Ave. (A1100), Austin, TX 78712, USA. Fax: +1-512-471-2957.
E-mail address: bchandra@utexas.edu (B. Chandrasekaran).

Abbreviations: ABR, auditory brainstem response; ANOVA, analysis of variance; EEG, electroencephalography; ERP, event-related potential; F0, fundamental frequency; FFR, frequency-following response; IC, inferior colliculus; MEG, magnetoencephalography; SSA, stimulus-specific adaptation; SVM, support vector machine.

https://doi.org/10.1016/j.neuroscience.2018.05.023
0306-4522/

debate. Per a central processing account, incoming sensory information is fully processed by both modalities at perceptual levels, the sensory information that gains prominence is determined at a central level of processing (Deutsch and Deutsch, 1963; Dux et al., 2006). However, a recent magnetoencephalography (MEG) study revealed substantial reduction in the early auditory cortical evoked activity to task-irrelevant auditory stimuli during high, relative to low, visual load (Molloy et al., 2015). This suggests that crossmodal influences can potentially be discerned even during the *encoding* stage of auditory processing. We directly test the hypothesis that visual load can influence the early sensory encoding of auditory signals by examining the frequency-following response (FFR), a scalp-recorded neurophonic component that faithfully captures phase-locked activity to periodic stimuli (Smith et al., 1975; Chandrasekaran and Kraus, 2010).

FFRs have been extensively used to index the fidelity of early, pre-attentive speech encoding in humans (Chandrasekaran and Kraus, 2010). The scalp-recorded FFRs derived from electroencephalography (EEG) are hypothesized to reflect activity primarily from subcortical auditory ensembles (Smith et al., 1975; Chandrasekaran and Kraus, 2010; Bidelman, 2015). However, there may also be cortical contributions to the FFR as well (Coffey et al., 2016). A recent proposal suggests that the FFRs reflect an integrated, dynamic interplay between pre-attentive cortical and subcortical circuitry (Kraus and White-Schwoch, 2015). This interplay is facilitated by ascending, as well as descending corticofugal pathways (Suga, 2008; Bajo and King, 2015; Malmierca et al., 2015).

Prior work has demonstrated that FFRs to speech stimuli are highly sensitive to stimulus context. Specifically, evidence has shown that sensory fidelity is enhanced for speech sounds presented in predictive contexts relative to less predictable contexts (Chandrasekaran et al., 2009; Slabu et al., 2012; Lau et al., 2016; Lehmann et al., 2016). The neural mechanism underlying context-dependent modulation of the FFRs is unclear. Two distinct mechanisms may be at play: (1) corticofugal modulation that selectively enhances the encoding of regularities in the signal via predictive processing; and (2) novelty detection, which reflects predominantly local stimulus-specific adaptation (SSA) (Chandrasekaran et al., 2014).

We examined the impact of manipulating visual perceptual load on the FFRs to speech signals under predictable and less predictable contexts. Native Mandarin Chinese speakers performed a visual search task of high or low perceptual load (Fig. 1B). On a random subset (50%) of trials, Mandarin lexical tones (Fig. 1A) were presented in a predictable context or a variable context (Fig. 1C). We recorded early cortical evoked activity, along with FFRs to the tones, and utilized a machine learning approach to decode speech category (Mandarin tone) information from the FFRs. We evaluate the extent to which decoding performance, which reflects the fidelity of stimulus encoding, is modulated by visual load and auditory stimulus context. Results reveal that, when the task-irrelevant speech stimuli were presented in variable contexts, the decodability of FFRs *increased* during higher visual load. But when speech stimuli were presented in predictable contexts, increasing visual load *reduced* the decodability of FFRs. We propose that a demanding visual task takes resources away from the auditory cortex, which 'releases' control from online predictive influences on lower-level sensory encoding. Under these conditions, we posit that stimulus encoding, as indexed by the FFRs, is geared toward the processing of less predictable (more novel) events. In contrast, in a less demanding visual task, the auditory cortex has available resources to enhance sensory tuning via predictive processing.

EXPERIMENTAL PROCEDURES

Participants

Twenty adult native speakers of Mandarin Chinese (9 females; 19–35 years old) took part in the study. All participants self-reported no previous history of hearing problems or neurological disorders. Participants underwent audiometric testing to ensure pure-tone thresholds $\leq$25-dB hearing level (HL) for octaves from 250 to 4000 Hz (less than 15 dB difference between the two ears) and had normal or corrected-to-normal vision. Each participant provided written, informed consent and received monetary compensation for their participation. The experimental protocol was approved by the Institutional Review Board at The University of Texas at Austin.

Stimuli and apparatus

Participants completed a visual search task in an acoustically shielded booth. The visual search task is similar to the task described in Experiment 1 in Molloy et al. (2015). The visual stimuli were presented on a zero latency VIEWPixx/EEG scanning LED-backlight LCD monitor (height: 29.1 cm, width: 52.2 cm; display resolution: 1920 * 1080; refresh rate: 120 Hz), placed $\sim$100 cm from the participants' eyes. As shown in Fig. 1B, the stimuli for the visual task consisted of six letters spaced about equally (subtending a viewing angle of $\sim$1.5°) around the center of the screen. The letters and the fixation cross were presented in white, and the display background was dark gray (RGB values: 77, 77, 77). One of the six letters was the target letter, X or Z (size = 0.55 $\times$ 0.45°) that occurred in equal probability. In the high-load condition (display *a* in Fig. 1B), letters K, W, V, N, and M (all with the same size as the target letters) served as the distracting items. In the low-load condition (display *b* in Fig. 1B), five smaller Os (size = 0.19 $\times$ 0.15°) were the distracting items. On each trial, we randomized the positions of the letters so that there was an equal probability for the target letter to appear in each of the six positions.

On a random 50% of trials, auditory stimuli were presented concurrently with the visual letter array (Fig. 1B) via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL) at 60-dB sound pressure level (SPL). The auditory stimuli were 100 ms

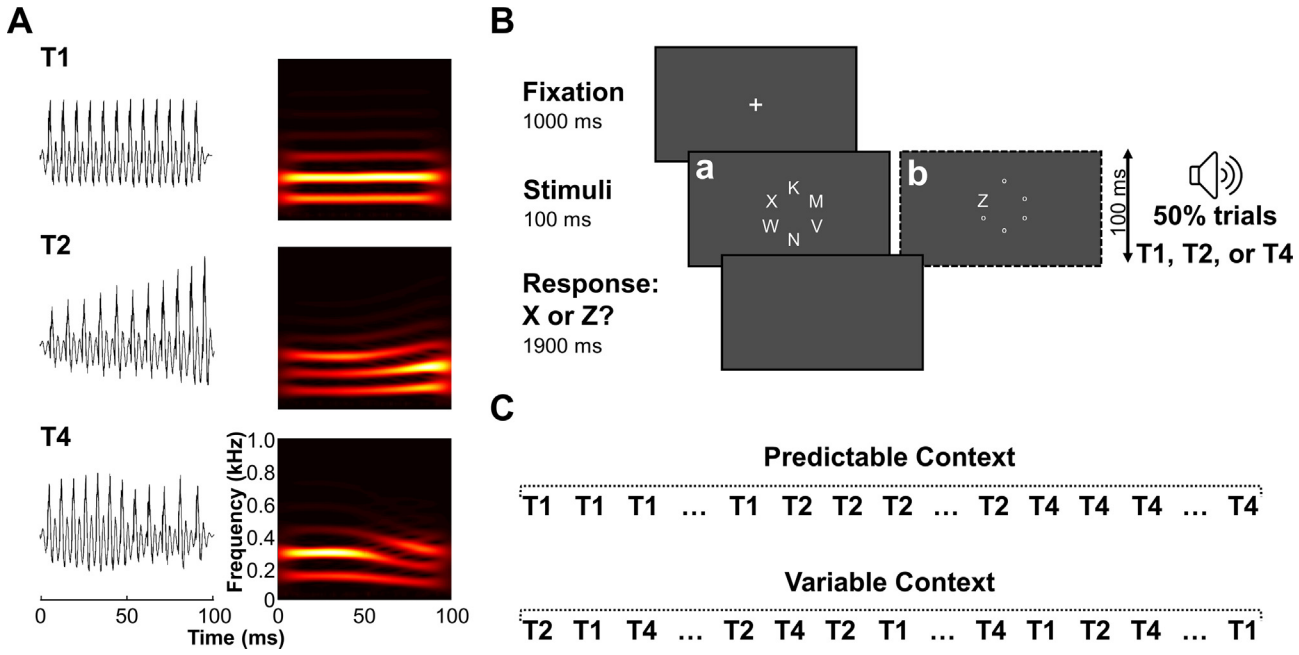

Fig. 1. Stimuli and task design. (A) Waveforms and spectrograms of the auditory stimuli, i.e., 100 ms Mandarin tones T1, T2, and T4. (B) Trial structure of the visual search task adapted from Molloy et al. (2015). Each trial began with a 1000-ms fixation cross at the center of the screen. Immediately after, a visual letter array of either (a) high or (b) low perceptual load was presented for 100 ms. On a random 50% of the trials, a 100-ms auditory stimulus (Mandarin tone T1, T2, or T4) was presented concurrently with the visual stimuli. In the remaining 50% trials, only the visual letter array was presented. After stimulus presentation, a blank screen was presented for a maximum of 1900 ms, during which participants were instructed to identify the visual target (X or Z) as quickly and accurately as possible. Once participants made a response, the task moved to the next trial. (C) The auditory stimuli were presented in either predictable (top) or variable (bottom) contexts. In the predictable auditory contexts, the tones were presented in blocks within which each tone was presented repetitively. In the variable auditory contexts, the tones were presented in a random order. In both contexts, an equal number of each of the three Mandarin tones were used.

long, diotically presented, linguistically relevant pitch patterns (Mandarin tones): T1, T2, and T4 (please see Fig. 1A for the waveforms and spectrograms). The three tones differ in fundamental frequency (F0) contours: T1 has a relatively flat F0 contour, T2 has a rising F0 contour, and T4 has a falling F0 contour. F0 contour is the primary acoustic cue for native Mandarin listeners to differentiate tones (Howie, 1976; Gandour, 1983). The tones were composed of the same syllable /i/, and were produced by a male native speaker of Mandarin Chinese, recorded at a sampling rate of 44.1 kHz. In pilot testing, native speakers ($n = 5$) reliably identified the tone categories with a high degree of accuracy ($> 90\%$).

Design and procedure

Participants completed the visual task in conditions of either high (display *a* in Fig. 1 B) or low (display *b* in Fig. 1 B) visual load, wherein the auditory stimuli were presented in either a predictable or a variable context. As shown in Fig. 1C, in the predictable context (top), the tones were presented in blocks within which each tone was presented repetitively. In the variable context (bottom), the tones were presented in a random order. In both auditory contexts, an equal number (96) of each of the three tones were used. Therefore, our study consisted of a two (visual load: high vs. low) $\times$ 2 (auditory stimulus context: predictable vs. variable) within-subject design. The four experimental conditions were divided into two sessions that were separated by seven to twelve days for 17 of 20 participants. For the remaining three participants, the session intervals were between 91 and 108 days due to scheduling conflicts. In each session, the auditory stimuli were presented in only one of the stimulus contexts (i.e., predictable or variable). Half the participants completed the session with predictable auditory contexts first, and the other half completed the session with variable auditory contexts first. In each session, there were 16 blocks (eight low visual load, eight high visual load) with 72 trials per block, each lasting about 3 min. Of the 72 trials in each block, 36 included auditory tones (12 per tone). The two visual load conditions were presented in alternating order and the order of the two load conditions was counterbalanced across participants.

The experiment was controlled with E-Prime 2.0.10 (Schneider et al., 2002). At the beginning of the study, participants were instructed that they may hear some sounds during the experiment. They were told to ignore the sounds and focus their attention on the visual task. Participants self-initiated each block. As illustrated in Fig. 1B, each trial began with a 1000-ms fixation cross at the center of the screen. Next, a visual letter array of either high (a) or low (b) load was presented for 100 ms. On 50% of the trials, a 100 ms Mandarin tone (T1, T2, or T4) was presented simultaneously with the visual display. In the remaining 50% of trials, only the visual letter array was presented. Immediately after the visual letter

array, a blank screen was presented, during which participants were required to identify the visual target as quickly and accurately as possible by pressing designated buttons with their right hand. After their response, the experiment immediately moved to the next trial. Participants had at most 1900 ms to respond. At the end of each block, visual task accuracy feedback was provided to encourage engagement. Between blocks, participants were allowed to take breaks when needed.

Electrophysiological data acquisition and preprocessing

Electrophysiological responses were continuously recorded with Ag/AgCl scalp electrodes placed at high forehead at the hairline (∼Fpz; active) referenced to linked mastoids (A1/A2), with another electrode on the mid-forehead as ground. Contact impedance was less than five kOhms for all electrodes. Responses were acquired at a sampling rate of 25 kHz using BrainVision PyCorder 1.0.7 (Brain Products, Gilching, Germany). The continuous EEG recordings were differentially bandpass filtered off-line from 1 to 30 Hz (12 dB/octave, zero phase-shift) and from 80 to 2500 Hz (12 dB/octave, zero phase-shift) to predominantly highlight cortical and subcortical sustained auditory electrophysiological responses, respectively (Musacchia et al., 2008; Bidelman and Alain, 2015). The EEG recordings were epoched into segments that are time locked to the auditory stimuli (cortical ERP: −100 to 300 ms; subcortical FFR: −40 to 150 ms), and to the visual stimuli in visual-only trials (cortical ERP: −100 to 300 ms). After baseline correcting each response to the mean voltage of the pre-stimulus region, trials with amplitudes exceeding a predefined range (cortical ERP: ± 100 μV; subcortical FFR: ± 50 μV) were rejected.

For auditory cortical ERPs, the artifact-free trials were averaged across all the three tones (T1, T2 and T4) for each condition, and downsampled from 25 kHz to 200 Hz. On average, at least 273.1 (*SD* = 33.34) out of the 288 possible trials (12 trials * 8 blocks * 3 tones) were used in each condition. The grand-average cortical ERPs across the four experimental conditions are shown in Fig. 2A. For visual cortical ERPs, the artifact-free trials were averaged for each condition, and downsampled from 25 kHz to 200 Hz. On average, 287.25 (*SD* = 1.45) out of the 288 possible trials (36 trials * 8 blocks) were used in each condition. The grand-average visual cortical ERPs across the four experimental conditions are shown in Fig. 2C.

To capture the FFRs, the artifact-free trials were averaged to produce a sample response to each tone at each condition and downsampled from 25 to 5 kHz. On average, at least 95.10 (*SD* = 1.07) out of the 96 possible trials (12 trials * 8 blocks) were used for any of the tones. Fig. 3 displays the grand-average subcortical FFRs to T2 across the four experimental conditions. On average, the signal-to-noise ratio (SNR), computed as the ratio of the root-mean-square amplitude of the post-stimulus region (10–110 ms) to that of the pre-stimulus region (−40 to 0 ms), is no lower than 1.24 (*SD* = 0.196) for any of the tones. The SNR was not significantly different across tone, visual load or auditory stimulus context (all *p*s > 0.129).

Analysis of cortical ERPs

Peak amplitude was measured for the N1 component of the auditory and visual cortical ERPs. The auditory N1 component reflects activity generated in auditory cortex and indexes early cortical processing of sounds (Näätänen and Picton, 1987). A prior study showed that the auditory M100, the magnetic equivalent of auditory N1, is reduced during high visual load, relative to low visual load (Molloy et al., 2015). Consistent with this study, we assessed the influence of visual load on auditory N1 responses. Note in that study visual load also modulated visual magnetic response M100. The EEG counterpart of visual M100 is visual P1 (Tobimatsu and Celesia, 2006). However, as displayed in Fig. 2C, we did not observe salient visual P1 component in the visual cortical ERPs. This is possibly because the recording site ∼Fpz is not optimal to pick up electrophysiological activity related to visual P1 responses (e.g., Alho et al., 1994; Vogel and Luck, 2000). We analyzed the visual N1 component in the visual cortical ERPs with the intent to demonstrate different visual load effects on the visual N1 response, relative to the auditory N1 response. We can thus make the inference that the auditory N1 responses in our data primarily reflected auditory activity. As illustrated in Fig. 2A and C, in each condition the N1 peak amplitude was taken as the maximal negative amplitude in a 60 ms time window around the N1 component of the grand-average response across the four conditions. The search for the 60-ms time window was conducted separately for auditory and visual cortical ERPs. The analysis was performed with custom MATLAB scripts (The MathWorks, Natick, MA).

Analysis of FFRs: Decoding information related to the Mandarin tones

Classification analysis was employed to examine the extent to which FFRs evoked by Mandarin tones contain relevant information to discriminate the tones in each of the four experimental conditions. We used a supervised machine learning algorithm (linear support vector machine; linear SVM; Christianini and Shawe-Taylor, 2000), implemented using the Scikit-learn library (Pedregosa et al., 2011) in python (http://scikit-learn.org/stable/). The linear SVM uses a "one-against-one" approach. Specifically, as there were three tones (T1, T2, and T4) in our experiment, linear SVM constructed three classifiers to test the FFR data from all the pairwise combinations of the three tones. The tone label with the highest probability was taken as the classified label. To ensure consistency across experimental conditions, we set the regulation parameter C at a fixed value of 0.1, while keeping other parameters at default values. The selection of this C value was based on our preliminary analysis with grid search to find the best parameter that maximizes tone classification performance.

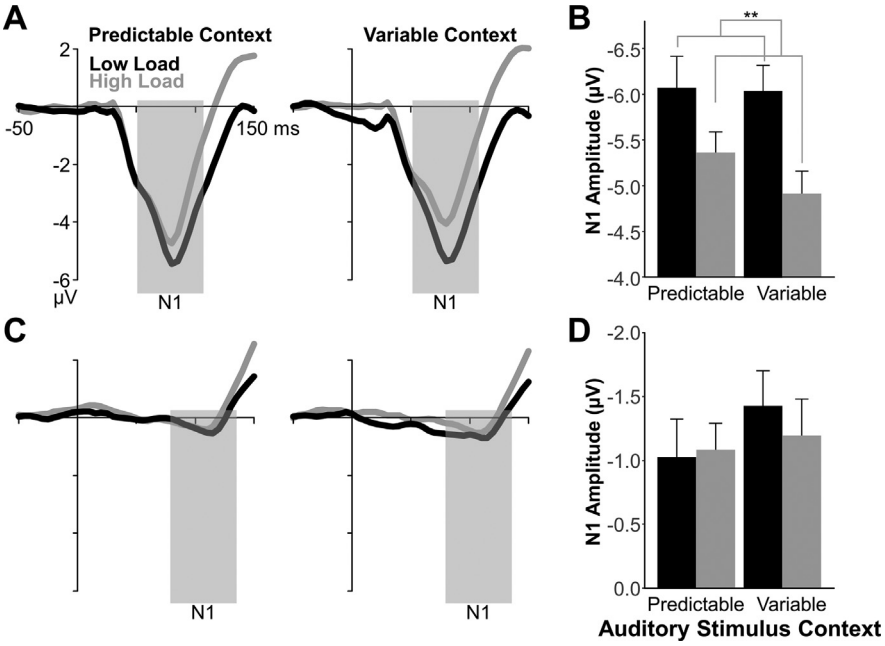

Fig. 2. Grand-average cortical evoked responses to the auditory stimuli (A and B), and to the visual stimuli in the visual-only trials (C and D) under low (black) or high (gray) visual perceptual load at predictable auditory contexts and variable auditory contexts. (A) Grand-average cortical evoked responses to the auditory stimuli. (B) Mean amplitude of N1 component for the auditory cortical evoked responses. (C) Grand-average cortical evoked responses to the visual stimuli in the visual-only trials. (D) Mean amplitude of N1 component for the visual cortical evoked responses. The shaded areas in A and C indicate the time window to find the N1 peak amplitude, which was defined as a 60-ms time window around the N1 component in the grand-average response across the four conditions. Error bars denote one standard error above the mean. $^{**}p < 0.01$.

Cross-validation strategy. To objectively evaluate the performance of the classifier, we used a four-fold cross-validation strategy with 5000 iterations. Each iteration began by randomizing the order of participants. Then, the FFR data were divided into four consecutive folds, with data from five unique participants in each fold. We trained the classifier with three of four folds (i.e., 15 out of 20 participants), and tested whether this training can generalize to the hold-out fold (i.e. the remaining five participants). We repeated this cross-validation four times until all the four folds had been tested against each other. Accuracy of each cross-validation was calculated, which reflected the percentage that the classifier correctly identified the tone labels of the FFR data. At each iteration, we calculated the decoding accuracy as the average accuracy across the four cross-validations. We estimated the decoding accuracy for each of the four experimental conditions at each iteration. Hence, 5000 decoding accuracy values were obtained to estimate the classifier's performance for each of the four experimental conditions.

Feature selection approaches. The latency of the FFR typically ranges 5–10 ms (Smith et al., 1975; Akhoun et al., 2008; Skoe et al., 2015), which is earlier than cortical evoked responses (Celesia et al., 1968; Moushegian et al., 1973). In line with our previous study (Xie et al., 2017), to account for onset delay reflecting subcortical (specifically midbrain) processes, we selected a region encompassing 10–110 ms (after stimulus onset) from each sample response, as representations of FFRs. For the first type of feature input, we used the raw amplitude value at each time point in the FFRs (10–110 ms post-stimulus; sampling rate of 5 kHz) (highlighted with orange rectangles in Fig. 3). In other words, the feature input consisted of 500 amplitude values from the FFRs. This type of feature input spans a frequency range of 80–2500 Hz. Next, to evaluate frequency-specific contribution to tone classification, we extracted spectrotemporal information spanning a narrow frequency band (80–180 Hz) that covers F0 range of all the three tones (~100 to ~140 Hz). We chose this frequency band because, as illustrated in the spectrograms of Fig. 3, much of the spectral energy in the FFRs is concentrated in the range of F0. We contrasted tone classification based on information from this frequency band (80–180 Hz) with that based on a higher frequency band (180–600 Hz). The higher frequency band encompasses the second through fourth harmonics (H2-H4) of all the tones. To derive these two frequency bands, we applied bandpass filtering (80–180 Hz and 180–600 Hz) to the original FFRs.

Statistical analysis. We applied the following analyses to the decoding performance of the three types of feature inputs separately. In the first analysis, we examined whether the obtained decoding accuracies were significantly above chance. To this end, we applied permutation tests ($n = 5000$) to test FFR decoding accuracies against a distribution of decoding accuracies obtained from randomly assigning the labels to the training data (i.e., null distribution). We first estimated the median of the 5000 decoding accuracies. We then estimated the p value using the formula: $p = (a + 1)/(n + 1)$ (Phipson and Smyth, 2010), where a is the number of decoding accuracies from the null distribution that exceeds the median of the FFR decoding accuracies, and n is the total number of decoding accuracies from the null distribution (i.e., 5000).

In the second analysis, we examined the effects of visual load and auditory stimulus context on the FFR decoding accuracies. To test the interaction between visual load and auditory stimulus context, we constructed a distribution of the difference between low and high visual load at each auditory stimulus context (i.e., predictable context or variable context). This was achieved by calculating the difference in decoding accuracy between low and high visual load at each

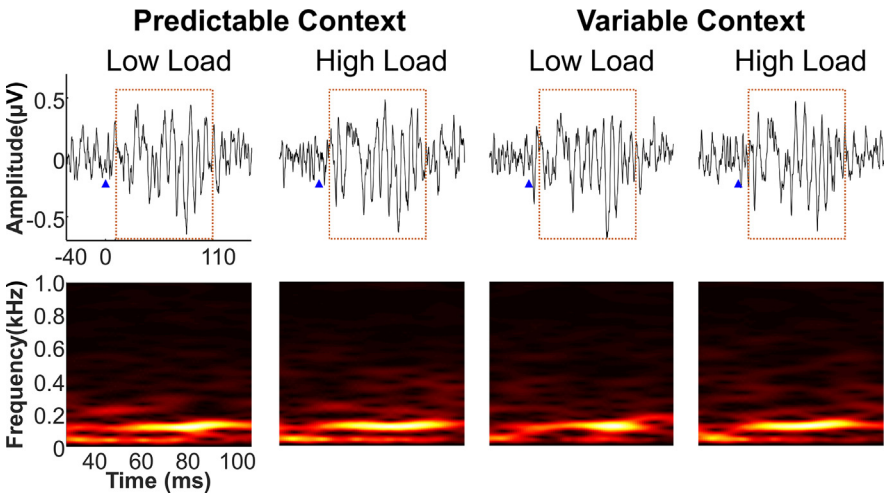

Fig. 3. Waveforms and spectrograms of the grand-average FFRs to Mandarin tone T2 in the four experimental conditions: predictable context + low load, predictable context + high load, variable context + low load, and variable context + high load. In the waveform plots, the blue triangles indicate the onset of the auditory stimulus. The dashed rectangles highlight FFRs from 10 to 110 ms (after stimuli onset) that were used as feature input for tone classification analysis. Specifically, the amplitude values (500 values) from this range were used for classification. The spectrograms correspond to FFRs at this range (i.e., 10–110 ms). Further, to evaluate frequency-specific contribution to tone classification, we also applied two bandpass filters (80–180 Hz and 180–600 Hz) to FFRs at range (i.e., 10–110 ms) to derive two new types of feature input: 80–180 Hz and 180–600 Hz. Amplitude values (500 values) from the two frequency bands were used for classification analysis, respectively. The frequency band 80–180 Hz covers F0 range of all the three tones (∼100 to ∼140 Hz). The frequency band 180–600 Hz encompasses the second through fourth harmonics (H2-H4) of all the tones. Note that, as shown in the spectrograms, much of the spectral energy in the FFR concentrates at the frequency range of 80–180 Hz, while limited spectral energy in the FFR was present at the frequency band of 180–600 Hz.

iteration, resulting in 5000 difference accuracy scores. We then estimated the median from the context condition with a higher value, and tested it against the distribution of difference scores from the context condition with a lower median value. We estimated the *p* values using the same procedures as described in the first analysis. If a significant interaction was found, we constructed four pairwise comparisons across the four experimental conditions. For each comparison, we estimated the median decoding accuracy from the condition with a higher median value, and tested it against the distribution of decoding accuracies from the condition with a lower median value. We estimated the *p* values using the same procedures as described in the first analysis. If no significant interaction between visual load and auditory stimulus context was found, we concatenated the decoding accuracies belonging to the same visual load condition or auditory stimulus context condition, and estimated the main effects of visual load and auditory stimulus context using the pairwise comparison method as described above.

Analysis of FFRs: Tracking of F0 contours in the Mandarin tones

To further understand the effects of visual load and auditory stimulus context on the neural encoding of the auditory stimuli (Mandarin tones) as reflected by the FFRs, we adopted the traditional approach to evaluate the fidelity of neural tracking of F0 contour in the Mandarin tones (e.g., Krishnan et al., 2004, 2005; Xie et al., 2017). Full details of the F0 tracking analysis are described in our previous study (Xie et al., 2017). We modified the parameters of this analysis to optimize application to the current study. Due to the constraint of using a behavioral task, the FFRs in the current study were averaged across far fewer number of trials (∼95 trials) relative to prior work with similar analysis (several hundreds to thousands; e.g., Krishnan et al., 2004, 2005; Xie et al., 2017). Hence, our results, compared to previous studies, are less robust to the influence of different sources of noise that may affect FFRs (Skoe and Kraus, 2010).

Extraction of F0 contours. We extracted the F0 contour from the FFRs (10–110 ms post-stimulus) using a sliding window (window size = 40 ms, step size = 1 ms) autocorrelation-based procedure (Boersma, 1993). The 40-ms slide window was applied to the time course of each FFR, producing a total of 60 overlapping bins. The autocorrelation function was applied each of the 60 bins to find the maximum (peak) autocorrelation value over a lag value of (1/180–1/80 m s), a range that encompasses the periods of the F0 contours for the three Mandarin tones. The peak autocorrelation value as well as the corresponding lag were recorded for each bin. The frequency of F0 at each bin was taken as the reciprocal of the lag at peak autocorrelation, resulting in a 60-point F0 contour. The same sliding window autocorrelation algorithm was applied to the evoking Mandarin tones to derive the respective stimulus F0 contour.

Evaluation of F0 tracking accuracy. We calculated two metrics to evaluate the robustness of the neural encoding of F0 contour as reflected by the FFRs: stimulus-to-response correlation and peak autocorrelation (e.g., Krishnan et al., 2005; Xie et al., 2017). Details for calculating the two metrics can be found in our previous study (Xie et al., 2017). In short, the stimulus-to-response correlation metric (ranging from 0 to 1) was computed as the normalized cross-correlation between F0 contours between the FFRs and the evoking stimulus. The peak autocorrelation metric (ranging from −1 to 1) was computed as the mean of the peak autocorrelation values across the 60 bins in the FFRs.

RESULTS

Behavioral: Performance in the visual search task

We employed a two-way repeated measures analysis of variance (ANOVA) to test the effects of visual load and

auditory stimulus context on performance in the visual search task. In this analysis, we only focused on the trials when the auditory stimuli were presented concurrently with the visual stimuli. Fig. 4 displays the accuracy rate and reaction time. For accuracy rate, we found a significant main effect of visual load [$F(1,19) = 54.956$, $p = 5.097 \times 10^{-7}$, $\eta_p^2 = 0.743$], indicating that the mean accuracy decreased in the high load (*mean* = 90.68%, *SD* = 6.92%) relative to the low-load condition (*mean* = 98.13%, *SD* = 2.36%) (Fig. 4A). The main effect of auditory stimulus context did not reach statistical significance [$F(1,19) = 0.54$, $p = 0.471$, $\eta_p^2 = 0.028$]. The interaction between visual load and auditory stimulus context was not significant [$F(1,19) = 0.019$, $p = 0.891$, $\eta_p^2 = 0.001$]. For task reaction time, we found a significant main effect of visual load [$F(1,19) = 261.46$, $p = 1.46 \times 10^{-12}$, $\eta_p^2 = 0.932$], indicating that mean reaction time increased in the high load (*mean* = 576.36 ms, *SD* = 82.13) relative to the low-load condition (*mean* = 425.2 ms, *SD* = 54.95) (Fig. 4B). The main effect of auditory stimulus context was also significant [$F(1,19) = 7.144$, $p = 0.015$, $\eta_p^2 = 0.273$], indicating slower reaction time in the variable contexts (*mean* = 512.43 ms, *SD* = 98.92) than the predictable contexts (*mean* = 489.13 ms, *SD* = 106.47) (Fig. 4B). The interaction between visual load and auditory stimulus context did not reach statistical significance [$F(1,19) = 0.127$, $p = 0.726$, $\eta_p^2 = 0.007$]. These findings suggest that when auditory stimuli were presented, a predictable auditory context facilitated performance on the visual task (i.e., faster reaction time) irrespective of the load of the visual task.

Auditory and visual cortical ERPs: N1 amplitude

In line with Malloy et al. (2015), we examined the effects of visual load and auditory stimulus context on the N1 amplitude from the cortical responses to the Mandarin tones, using a two-way repeated measures ANOVA. The mean N1 amplitude of the auditory cortical EPRs are shown in Fig. 2B. We found a significant main effect of visual load [$F(1,19) = 9.946$, $p = 5.228 \times 10^{-3}$, $\eta_p^2 = 0.344$], indicating that the mean N1 amplitude decreased in the high load (*mean* = −5.139 μV, *SD* = 2.836) relative to the low-load condition (*mean* = -6.054 μV, *SD* = 3.345). The main effect of auditory stimulus context did not reach statistical significance [$F(1,19) = 0.537$, $p = 0.473$, $\eta_p^2 = 0.0275$]. The interaction between visual load and auditory stimulus context was not significant [$F(1,19) = 1.169$, $p = 0.293$, $\eta_p^2 = 0.058$].

Note that the Mandarin tones were presented concurrently with visual stimuli with different physical properties in the two load conditions (see Fig. 1B). Hence, one possibility is that the load-related differences in N1 amplitude of the cortical response to Mandarin tones predominantly reflect differences in the visual evoked responses. Such possibility can be refuted because, based on Molloy et al. (2015), we would predict increased cortical responsivity for high visual load relative to low visual load for the visual evoked responses. Further, we directly examined the effects of visual load and auditory stimulus context on the N1 amplitude from visual cortical responses in trials that included only the visual stimuli (i.e., visual "alone" trials). The mean N1 amplitude of visual cortical EPRs is shown in Fig. 2D. We did not find significant main effect of visual load [$F(1,19) = 0.153$, $p = 0.7$, $\eta_p^2 = 0.008$] or auditory stimulus context [$F(1,19) = 0.455$, $p = 0.508$, $\eta_p^2 = 0.023$], or significant interaction between visual load and auditory stimulus context [$F(1,19) = 0.861$, $p = 0.365$, $\eta_p^2 = 0.043$]. These results again suggest that the load-related differences in N1 amplitude of the auditory cortical response predominantly reflect differences in auditory activity.

FFRs: Decoding information related to the Mandarin tones

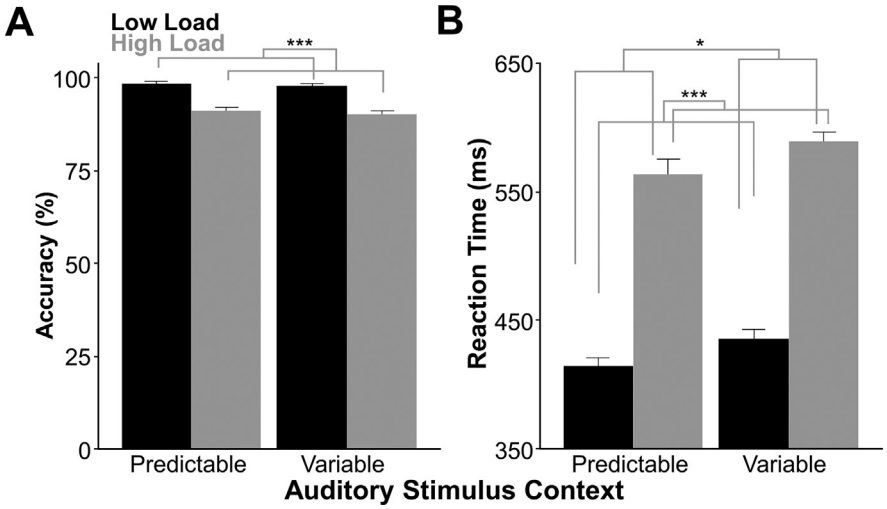

Fig. 4. Mean accuracy rate (A) and reaction time (B) for identifying the target (X or Z) in the visual search task. Data are presented only for trials with auditory stimuli. The auditory stimuli (Mandarin tone T1, T2, or T4) were presented in predictable or variable contexts. In each auditory stimulus context, the visual stimuli were of low (black) or high (gray) visual perceptual load. $^{*}p < 0.05$, $^{***}p < 0.001$.

Feature input of 80–2500 Hz. We first examined the extent to which the FFR decoding accuracies were significantly above chance. Permutations tests showed that decoding accuracies (left panel in Fig. 5) were significantly above chance level (indicated by the blue triangles) across the four experimental conditions (all *p*s = 1.9 996 × 10^{-4}). Next, we examined the effects of visual load and auditory stimulus context on FFR decoding accuracies. There was a significant interaction between visual load and auditory stimulus context ($p = 1.999\,6 \times 10^{-4}$). Follow-up pairwise comparisons showed that, as displayed in the left panel of Fig. 5, in the predictable auditory context, decoding accuracies were

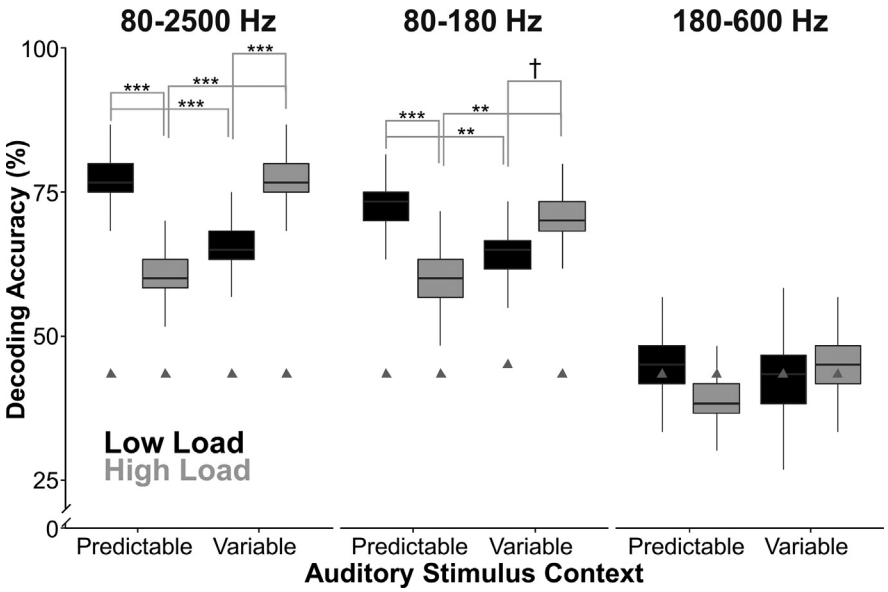

Fig. 5. Boxplots of the accuracies to decode information related to the Mandarin tones from the FFRs. The classification analysis used features (500 amplitude values) from *post*-stimulus region (10–100 ms) which covers the frequency range 80–2500 Hz (left), 80–180 Hz (middle), and 180–600 Hz (right), respectively. The auditory stimuli (Mandarin tone T1, T2, or T4) were presented in either predictable or variable contexts, wherein participants performed a visual task of either low (black) or high (gray) visual load. There were 5000 iterations in each classification analysis, yielding 5000 decoding accuracies. The boxes and the horizontal line inside show the quartiles (1st to 3rd quartile) and the median, respectively. The whiskers denote 1.5 times the interquartile range. Outliers, defined as cases with values outside the 1.5 interquartile range, were not displayed here but were included for statistical analysis. The blue triangles indicate the 95th percentile of decoding accuracies from permutation tests. $^{\dagger}p < 0.1$, $^{**}p < 0.01$, $^{***}p < 0.001$.

significantly *lower* for the high visual load condition relative to the low visual load condition ($p = 1.9996 \times 10^{-4}$, uncorrected). However, in the variable auditory context, decoding accuracies were significantly *higher* for the high-load condition than the low-load condition ($p = 9.998 \times 10^{-4}$, uncorrected). In the low-load condition, decoding accuracies for the predictable auditory context were significantly *higher* than for the variable auditory context ($p = 9.998 \times 10^{-4}$, uncorrected). But in the high visual load condition, decoding accuracies for the predictable auditory context were significantly *lower* than for the variable auditory context ($p = 1.9996 \times 10^{-4}$, uncorrected).

Feature input of 80–180 Hz. We first examined the extent to which the FFR decoding accuracies were significantly above chance. Permutations tests showed that decoding accuracies (middle panel in Fig. 5) were also significantly above chance level (indicated by the blue triangles) across the four experimental conditions (all ps $= 1.9996 \times 10^{-4}$). Then, we tested the effects of visual load and auditory stimulus context on FFR decoding accuracies. As displayed in the middle panel of Fig. 5, the patterns of decoding accuracies were similar to those found for feature input of 80–2500 Hz. Statistically, there was a significant interaction between visual load and auditory stimulus context ($p = 3.9992 \times 10^{-4}$). Follow-up pairwise comparisons showed that in the predictable auditory context decoding accuracies were significantly *lower* for the high visual load condition, relative to the low visual load condition ($p = 1.9996 \times 10^{-4}$, uncorrected). However, in the variable auditory context decoding accuracies were marginally significantly *higher* for the high-load condition than the low-load condition ($p = 0.06879$, uncorrected). In the low-load condition, decoding accuracies for the predictable auditory context were significantly *higher* than for the variable auditory context ($p = 6.3988 \times 10^{-3}$, uncorrected). But in the high visual load condition, decoding accuracies for the predictable auditory context were significantly *lower* than for the variable auditory context ($p = 2.599\,5 \times 10^{-3}$, uncorrected).

Feature input of 180–600 Hz. We first examined the extent to which the FFR decoding accuracies were significantly above chance. Permutations tests showed that decoding accuracies (right panel in Fig. 5) were significantly above chance level (indicated by the blue triangles) for three of the four experimental conditions (predictable context + low load: $p = 0.025795$; variable context + low load: $p = 0.047391$; variable context + high load: $p = 0.025795$). Decoding accuracies for the remaining condition (predictable context + high load) were not significantly above chance level ($p = 0.14917$). Then, we tested the effects of visual load and auditory stimulus context on FFR decoding accuracies. There was no significant interaction between visual load and auditory stimulus context ($p = 0.13937$), or significant main effect of visual load ($p = 0.33297$) or auditory stimulus context ($p = 0.33487$).

FFRs: Tracking of F0 contours in the Mandarin tones

We employed a three-way repeated measures ANOVA to test the effects of visual load and stimulus context on the neural tracking of F0 contours in the Mandarin tones. In this analysis, visual load (high vs. low), auditory stimulus context (predictable vs. variable), and tone (T1, T2, and T4) were included as with-subject factors. We converted stimulus-to-response correlation and peak autocorrelation into Fisher's *Z* scores to improve the normality of the data and used the converted Z scores for statistical analyses (Wong et al. 2007; Xie et al., 2017).

For the stimulus-to response metric, the main effect of auditory stimulus context was marginally significant [$F(1,19) = 3.97$, $p = 0.061$, $\eta_p^2 = 0.173$], indicating that the mean stimulus-to response was higher in the predictable auditory context (*mean* = 0.675, *SD* = 0.249) relative to the variable condition (*mean* = 0.627, *SD* = 0.234). The interaction between visual load and

auditory stimulus context was marginally significant [$F(1,19) = 4.283$, $p = 0.052$, $\eta_p^2 = 0.184$]. Simple effect analysis revealed that, the mean stimulus-to response was higher in the low-load condition (*mean* = 0.72, *SD* = 0.262) relative to the high-load condition (*mean* = 0.629, *SD* = 0.229) for the predictable auditory context [$t(59) = -2.191$, $p = 0.0324$, uncorrected], but not for the variable auditory context [low load: *mean* = 0.626, *SD* = 0.257; high load: *mean* = 0.628, *SD* = 0.212; $t(59) = 0.0369$, $p = 0.971$, uncorrected]. The main effects of visual load or tone, or two-way or three-way interaction between visual load, auditory stimulus context and tone did not reach significance (all $p > 0.093$, η_p^2 ranging from 0.015 to 0.123). For the peak autocorrection metric, none of the main effects, two-way or three-way interaction between visual load, auditory stimulus context and tone were significant (all $p > 0.109$, η_p^2 ranging from 0.013 to 0.12). It is important to note that due to the constraints of using a behavioral task, the number of FFR trials (~95 trials per tone) is extremely low relative to typical studies examining the FFRs (e.g., Krishnan et al., 2004, 2005; Xie et al., 2017). Despite this, we see trends in the same direction as the machine learning classification metrics.

DISCUSSION

We examined the extent to which visual perceptual load modulates the early sensory encoding of speech signals. Our results demonstrate that the early sensory encoding of speech signals, as indexed by the FFRs, was modulated by the level of perceptual load in the visual task, as well as the context in which the task-irrelevant speech stimuli were presented. When irrelevant speech stimuli were presented in predictable contexts, increasing visual load *reduced* the decodability of FFRs. However, an opposite pattern was observed when the speech stimuli were presented in variable contexts, such that the decodability of FFRs *increased* with higher visual load. These findings suggest that focusing attention on a visual task of high perceptual demand influences early auditory encoding, but in a context-dependent manner. The direction of visual attentional influence is highly contingent on the predictability of the incoming auditory stream.

Load theory posits that visual and auditory processing share central, capacity-limited neural resources, and the depletion of resources by one modality diminishes available resources to the other modality (Lavie, 2005; Klemen et al., 2009). Hence, increasing perceptual load on a visual task would lead to reduced availability of neural resources for the processing of task-irrelevant auditory stimuli (Macdonald and Lavie, 2011; Raveh and Lavie, 2015). A recent study focusing on auditory cortical processing demonstrated that higher visual load is associated with decreased auditory cortical responses to irrelevant auditory stimuli (Molloy et al., 2015). Similarly, we demonstrate that the early auditory cortical activity, as indexed by the N1 response, is reduced during high visual load, relative to low visual load. Critically, we demonstrate that the load modulation of auditory processing can be evidenced even at the earliest levels of sensory processing involving stimulus encoding, as indexed by the FFRs.

Notably, our findings suggest that additional mechanisms are at play in mediating the impact of visual load on early auditory processing. Our results likely reflect a complex interaction between top-down and bottom-up processes in mediating the auditory responsivity to task-irrelevant stimuli. Animal studies suggest that auditory cortical modulation can fine-tune the encoding of auditory signals in subcortical nuclei (Yan and Suga, 1996; Suga et al., 1997). Such corticofugal modulatory influence is argued to be important for the selection of regularities in the stimulus stream. Indeed, a prominent role of the auditory cortex is in predictive processing, i.e., making continuous predictions based on prior experience in order to enhance bottom-up signals. Enhanced fidelity of the FFR in predictive contexts may reflect cortical tuning to enhance the encoding of predictable regularities in the incoming stimulus stream. In addition, animal models have also revealed that subcortical neurons, especially in the inferior colliculus (IC), are highly sensitive to novelty and adapt locally in a stimulus-specific manner. This leads to a decrease in responsivity to repetitive stimuli and heightened responsivity to less predictable stimuli (Malmierca et al., 2009; Anderson and Malmierca, 2013). This form of SSA is predominantly a locally-generated process (Anderson and Malmierca, 2013; Duque and Malmierca, 2015) that gears toward novelty detection.

Based on our results, we posit that there is a constant push–pull between auditory cortical modulation (reflecting predictive processes) and locally driven processes like SSA (reflecting novelty detection) in mediating sensory encoding, as indexed by the FFR. This argument is supported by an animal study demonstrating that IC neurons exhibiting SSA also receive feedback projections from auditory cortex to the IC (Ayala et al., 2015), which provides the infrastructure for dynamic auditory cortical *control* over local subcortical processes. We posit that the low-load visual task leaves enough resources available for auditory processing, such that the stronger involvement of auditory cortical top-down modulation overrides novelty detection in favor of enhancing predictable elements. Hence, the dominance of auditory cortical control may sharpen the encoding of *predictable* stimuli at early subcortical levels of processing. When the task involves higher visual perceptual load, all or most of the shared neural resources are consumed, and little or none is left for mediating top-down auditory control. This leads to diminished auditory cortical activity (Molloy et al., 2015), and weaker auditory cortical top-down control of the IC (Zhang and Suga, 1997; Bajo et al., 2010; Anderson and Malmierca, 2013). However, locally generated processes like SSA are still preserved (Anderson and Malmierca, 2013), and may in fact become more dominant. Dominance of local processes may be the norm during sleep, for example, where there is an important benefit for gearing the system toward novelty detection (e.g., waking up to threat). The combined

effects of decreased auditory cortical control and preserved SSA at the IC may lead to less robust subcortical encoding of predictable stimuli less robust, but enhanced subcortical encoding of variable, less predictable stimuli.

Based on the load theory, it is assumed that visual processing is prioritized in our tasks (Lavie, 2005; Macdonald and Lavie, 2011; Raveh and Lavie, 2015). That is, because participants are instructed to perform the visual search task and ignore the incoming auditory stimuli, attentional resources will first be allocated to visual processing. If there are attentional resources left, these available resources will then be used for auditory processing. In other words, the visual task would not be affected by the presence of the auditory stimuli. This assumption is in direct contrast to our behavioral findings, where the predictability of the auditory stimulus stream influenced performance on the visual task, such that visual targets presented concurrently with unpredictable auditory stimuli (variable context) were associated with overall prolonged reaction time. Therefore, the assumption that visual processing is prioritized may be violated when stimuli from the auditory modality are variably presented. This may be because variable, unpredictable auditory stimuli may be more distracting and require more resources to process, compared to predictable auditory stimuli (Southwell et al., 2017). This may explain why we observed a different pattern of visual load effect on the FFRs to speech stimuli between variable and predictable auditory contexts.

Importantly, we investigated the relevance of features of the FFR for its decodability under different visual load and auditory predictability conditions. Our results indicate that the F0 of the FFRs might be the feature changing as a function of visual load and auditory predictability. First, decoding accuracies of FFRs were well above chance and influenced by visual load and auditory predictability for FFRs with spectrotemporal information covers the F0 range of all the three tones (~100 to ~140 Hz), but not when the FFRs were filtered above the F0 range (180–600 Hz). The current finding of frequency-specific modulation by visual load is in line with a recent study demonstrating that attention modulated FFRs to stimuli with modulation rates at ~ 100 Hz, but not to stimuli with modulation rates at above 200 Hz (Holmes et al., 2017). Further, we directly examined the neural encoding of F0 contours as indexed by the FFRs. Partly consistent with the decoding results, we found a more faithful encoding of F0 contours (higher stimulus-to-response correlation) in the low-load condition than the high-load condition for the predictable auditory context, but not for the variable auditory context.

In a recent MEG study, Coffey et al. (2016) demonstrated a contribution from auditory cortex to the F0 of the FFR close to 100 Hz, in addition to contributions from subcortical nuclei. This raises the possibility that the FFRs recorded in the current study reflect auditory cortical contribution, given that our auditory stimuli have F0s from ~100 to ~140 Hz and substantial energy in these regions was found in the corresponding FFRs. However, the possibility that we are examining cortical encoding is unlikely to explain our main findings. First, FFR derived from scalp-recorded EEG (as in the current study) and from MEG likely reflect different source contributions (Cohen and Cuffin, 1983; Goldenholz et al., 2009; Ahlfors et al., 2010). Interpretations regarding sources from MEG cannot be directly applied to EEG. A recent study (Bidelman, 2015) used a stimulus with a similar F0 (88–120 Hz) to our study and examined different source contributions of the FFR using multichannel scalp-recorded EEG. This study indicated that the sources of FFR are consistent with generators in the IC. Second, King et al. (2016) found that FFRs at 85–145 Hz (similar to the F0 range of our stimuli) has a latency about 8–9 ms, suggesting sources in the rostral brainstem or IC (Møller and Jannetta, 1982). Third, in the present study we found a consistent effect of visual load (reduced N1 amplitude for high vs. low load) on auditory cortical responses across auditory stimulus contexts (Fig. 2B). However, the data-driven decoding results do not reflect a simple effect of visual load. We observed an interaction between visual load and auditory stimulus context (see left and middle panels in Fig. 5). Based on our results, we suggest that modulating visual load and auditory stimulus context can be utilized as an experimental strategy to evaluate the relative contribution of multiple top-down and bottom-up processes that influence speech encoding.

The extent to which cross-modal attention modulates human FFR is the subject of intense debate. In a recent review, Varghese et al. (2015) found no effects of visual attention on the FFR. However, Galbraith and colleagues (2003) demonstrated that visual attention decreased the amplitude of FFRs to repetitive tones. Hairston et al. (2013) also showed that visual attention reduced the robustness of FFRs to repetitive tones. These mixed findings may reflect variation in the degree to which listeners disengage their attention from incoming auditory stimuli, and/or on the extent of overlap between auditory and visual stimuli presentation. For example, Sörqvist et al. (2012) examined the effects of attention on subcortical auditory responses, as reflected by wave V of auditory brainstem response (ABR). Wave V of the ABR is a transient counterpart of the FFR, and is also thought to primarily originate from the IC (Møller and Jannetta, 1985). They found that, the wave V of ABR remained unchanged from an active listening condition to a condition where listeners performed a visual task of low attentional demand. A significant reduction in the wave V amplitude was only observed when the attentional demand of the visual task increased. Consistent with these findings, the current study systematically manipulated the perceptual load of the visual task, and provided evidence for the modulation of the FFR by visual attention. Further, task paradigms utilized in previous studies did not require listeners to consistently maintain attention on a simultaneously presented visual task, leaving opportunities for attentional capture by the auditory stimuli. To manipulate visual attention more rigorously, future studies may adapt paradigms similar to the current study, wherein participants are required to consistently focus attention to brief visual stimuli that *coincide* with the auditory stimuli.

In conclusion, our data provide important insights into the mechanisms of multisensory processing. When

the brain is overloaded with sensory information from various modalities, the competition for central, capacity-limited perceptual resources among the modalities impacts early encoding of sensory inputs in the task-irrelevant modality. Critically, this influence is dependent on the predictability of the incoming stimulus stream, a reflection of a likely push–pull dynamic between predictive processes and novelty detection within the auditory system.

DECLARATIONS OF INTEREST

None.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and Other Communication Disorders-National Institutes of Health (Grants R01DC013315 and R01DC015504 to B.C). We thank Jacie McHaney, Cat Han, Rachel Tessmer and the rest of the SoundBrain laboratory research assistants for their assistance with participant recruitment and data collection. We thank Gangyi Feng and Han Gyol Yi for their help with data analysis and manuscript preparation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

Ahlfors SP, Han J, Belliveau JW, Hämäläinen MS (2010) Sensitivity of MEG and EEG to source orientation. Brain Topogr 23:227–232.

Akhoun I, Gallégo S, Moulin A, Ménard M, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H (2008) The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme/ba/in normal-hearing adults. Clin Neurophysiol 119:922–933.

Alho K, Woods DL, Algazi A (1994) Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. Psychophysiology 31:469–479.

Anderson L, Malmierca M (2013) The effect of auditory cortex deactivation on stimulus-specific adaptation in the inferior colliculus of the rat. Eur J Neurosci 37:52–62.

Ayala YA, Udeh A, Dutta K, Bishop D, Malmierca MS, Oliver DL (2015) Differences in the strength of cortical and brainstem inputs to SSA and non-SSA neurons in the inferior colliculus. Sci Rep 5:10383.

Bajo VM, King AJ (2015) Cortical modulation of auditory processing in the midbrain. Inferior Colliculus Microcircuits 134.

Bajo VM, Nodal FR, Moore DR, King AJ (2010) The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat Neurosci 13:253–260.

Bidelman GM (2015) Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. Hear Res 323:68–80.

Bidelman GM, Alain C (2015) Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. J Neurosci 35:1240–1249.

Boersma P (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: Proceedings of the institute of phonetic sciences, vol. 17, pp 97–110: Amsterdam.

Celesia GG, Broughton RJ, Rasmussen T, Branch C (1968) Auditory evoked responses from the exposed human cortex. Electroencephalogr Clin Neurophysiol 24:458–465.

Chandrasekaran B, Hornickel J, Skoe E, Nicol T, 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.

Chandrasekaran B, Kraus N (2010) The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 47:236–246.

Chandrasekaran B, Skoe E, Kraus N (2014) An integrative model of subcortical auditory plasticity. Brain Topogr 27:539–552.

Christianini N, Shawe-Taylor J (2000) Support vector machines. Cambridge, UK: Cambridge University Press. 93:935–948.

Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. Nat Commun 7.

Cohen D, Cuffin BN (1983) Demonstration of useful differences between magnetoencephalogram and electroencephalogram. Electroencephalogr Clin Neurophysiol 56:38–51.

Deutsch JA, Deutsch D (1963) Attention: Some theoretical considerations. Psychol Rev 70:80.

Duque D, Malmierca MS (2015) Stimulus-specific adaptation in the inferior colliculus of the mouse: anesthesia and spontaneous activity effects. Brain Struct Funct 220:3385–3398.

Dux PE, Ivanoff J, Asplund CL, Marois R (2006) Isolation of a central bottleneck of information processing with time-resolved fMRI. Neuron 52:1109–1120.

Galbraith GC, Olfman DM, Huffman TM (2003) Selective attention affects human brain stem frequency-following response. Neuroreport 14:735–738.

Gandour J (1983) Tone perception in far eastern-languages. J Phon 11:149–175.

Goldenholz DM, Ahlfors SP, Hämäläinen MS, Sharon D, Ishitobi M, Vaina LM, Stufflebeam SM (2009) Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography. Hum Brain Mapp 30:1077–1086.

Hairston WD, Letowski TR, McDowell K (2013) Task-related suppression of the brainstem frequency following response. PLoS One 8 e55215.

Holmes E, Purcell DW, Carlyon RP, Gockel HE, Johnsrude IS (2017) Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. J Assoc Res Otolaryngol:1–15.

Howie JM (1976) Acoustical studies of Mandarin vowels and tones. Cambridge University Press.

King A, Hopkins K, Plack CJ (2016) Differential group delay of the frequency following response measured vertically and horizontally. J Assoc Res Otolaryngol 17:133–143.

Klemen J, Büchel C, Rose M (2009) Perceptual load interacts with stimulus processing across sensory modalities. Eur J Neurosci 29:2426–2434.

Kraus N, White-Schwoch T (2015) Unraveling the biology of auditory learning: a cognitive–sensorimotor–reward framework. Trends Cogn Sci 19:642–654.

Krishnan A, Xu Y, Gandour J, Cariani P (2005) Encoding of pitch in the human brainstem is sensitive to language experience. Cogn Brain Res 25:161–168.

Krishnan A, Xu Y, Gandour JT, Cariani PA (2004) Human frequency-following response: representation of pitch contours in Chinese tones. Hear Res 189:1–12.

Lau JC, Wong PC, Chandrasekaran B (2016) Context-dependent plasticity in the subcortical encoding of linguistic pitch patterns. J Neurophysiol. https://doi.org/10.1152/jn.00656.2016.

Lavie N (2005) Distracted and confused?: Selective attention under load. Trends Cogn Sci 9:75–82.

Lehmann A, Arias DJ, Schönwiesner M (2016) Tracing the neural basis of auditory entrainment. Neuroscience 337:306–314.

Macdonald JS, Lavie N (2011) Visual perceptual load induces inattentional deafness. Atten Percept Psychophys 73:1780–1789.

Malmierca MS, Anderson LA, Antunes FM (2015) The cortical modulation of stimulus-specific adaptation in the auditory midbrain and thalamus: a potential neuronal correlate for predictive coding. Front Syst Neurosci 9:19.

Malmierca MS, Cristaudo S, Pérez-González D, Covey E (2009) Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. J Neurosci 29:5483–5493.

Møller AR, Jannetta P (1985) Neural generators of the auditory brainstem response. The auditory brainstem response 13–31.

Møller AR, Jannetta PJ (1982) Evoked potentials from the inferior colliculus in man. Electroencephalogr Clin Neurophysiol 53:612–620.

Molloy K, Griffiths TD, Chait M, Lavie N (2015) Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses. J Neurosci 35:16046–16054.

Moushegian G, Rupert AL, Stillman RD (1973) Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr Clin Neurophysiol 35:665–667.

Musacchia G, Strait D, Kraus N (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear Res 241: 34–42.

Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375–425.

Nelken I, Ulanovsky N (2007) Mismatch negativity and stimulus-specific adaptation in animal models. J Psychophysiol 21:214–223.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830.

Phipson B, Smyth GK (2010) Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. Stat Appl Genet Mol Biol 9 39.

Raveh D, Lavie N (2015) Load-induced inattentional deafness. Atten Percept Psychophys 77:483–492.

Schneider W, Eschman A, Zuccolotto A (2002) E-Prime: User's guide: Psychology Software Incorporated.

Skoe E, Kraus N (2010) Auditory brainstem response to complex sounds: a tutorial. Ear Hear 31:302.

Skoe E, Krizman J, Anderson S, Kraus N (2015) Stability and plasticity of auditory brainstem function across the lifespan. Cereb Cortex 25:1415–1426.

Slabu L, Grimm S, Escera C (2012) Novelty detection in the human auditory brainstem. J Neurosci 32:1447–1452.

Smith JC, Marsh JT, Brown WS (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr Clin Neurophysiol 39:465–472.

Sörqvist P, Stenfelt S, Rönnberg J (2012) Working memory capacity and visual–verbal cognitive load modulate auditory–sensory gating in the brainstem: Toward a unified view of attention. J Cognit Neurosci 24:2147–2154.

Southwell R, Baumann A, Gal C, Barascud N, Friston K, Chait M (2017) Is predictability salient? A study of attentional capture by auditory patterns. Phil Trans R Soc B 372 20160105.

Stefanics G, Kremláček J, Czigler I (2014) Visual mismatch negativity: a predictive coding view. Front Human Neurosci 8.

Suga N (2008) Role of corticofugal feedback in hearing. J Comp Physiol A 194:169–183.

Suga N, Yan J, Zhang Y (1997) Cortical maps for hearing and egocentric selection for self-organization. Trends Cogn Sci 1:13–20.

Tobimatsu S, Celesia GG (2006) Studies of human visual pathophysiology with visual evoked potentials. Clin Neurophysiol 117:1414–1433.

Varghese L, Bharadwaj HM, Shinn-Cunningham BG (2015) Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. Brain Res 1626:146–164.

Vogel EK, Luck SJ (2000) The visual N1 component as an index of a discrimination process. Psychophysiology 37:190–203.

Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn Sci 13:532–540.

Xie Z, Reetzke R, Chandrasekaran B (2017) Stability and plasticity in neural encoding of linguistically relevant pitch patterns. J Neurophysiol 117:1407–1422.

Yan J, Suga N (1996) Corticofugal modulation of time-domain processing of biosonar information in bats. Science 273:1100.

Zhang Y, Suga N (1997) Corticofugal amplification of subcortical responses to single tone stimuli in the mustached bat. J Neurophysiol 78:3489–3492.

(Received 28 November 2017, Accepted 16 May 2018)
(Available online 24 May 2018)