Neural correlates of auditory enhancement in humans

Lei Feng[1], Anahita H. Mehta[1] and Andrew J. Oxenham

Department of Psychology, University of Minnesota, 75 East River Parkway, Minneapolis, Minnesota 55455, USA

fengl@umn.edu

mehta@umn.edu

oxenham@umn.edu

Running title: Neural correlates of auditory enhancement

[1] These authors contributed equally to this work.

ABSTRACT

A target sound embedded within a background sound becomes perceptually more salient if the background is presented first by itself. This phenomenon, known as auditory enhancement, reflects a general principle of contrast enhancement, and may help in the detection of new acoustic events in the environment and in establishing the perceptual constancy of speech and other biologically relevant sounds under varying acoustic conditions. Surprisingly, no neural correlates of this important phenomenon have been reported in humans. Here we used the auditory steady state response (ASSR) to determine whether the neural response to the target is amplified under conditions of enhancement. We used a double-modulation paradigm, involving the simultaneous amplitude modulation of a tone with two modulation frequencies, to distinguish cortical from subcortical contributions to this phenomenon. Robust phase-locked neural responses to both the target and masker were identified at both cortical and subcortical levels. Consistent with perceptual results, the response to the target tone embedded in the simultaneous maskers increased in the presence of the precursor, whereas the response to the masker components remained constant across conditions. The quantitative pattern of results suggest that the enhancement effects emerge at a subcortical level but are further enhanced within the auditory cortex.

Keywords: auditory perception, contrast enhancement, perceptual invariance, EEG, ASSR

INTRODUCTION

A target embedded within a background sound can "pop out" perceptually if the background masker is presented by itself first. This phenomenon, commonly termed auditory enhancement (Viemeister, 1980; Viemeister and Bacon, 1982), has been demonstrated and quantified in many psychophysical studies. For instance, thresholds for detecting a target tone within a simultaneous masker can be improved (decreased) by presenting a copy of the masker in the form of a precursor (Viemeister, 1980). A precursor can also increase the effectiveness of a target tone in masking a subsequent probe tone, suggesting that the target's neural representation has been amplified (Viemeister and Bacon, 1982). When the target is presented well above its detection threshold, its perceived loudness can be increased by a precursor (Wang and Oxenham, 2016), and it can become sufficiently salient to be perceived as a separate entity with a distinct pitch (Hartmann and Goupell, 2006; Erviti et al., 2011; Byrne et al., 2013; Demany et al., 2013; Feng and Oxenham, 2015).

Auditory enhancement may reflect processes in the auditory system that aim to adapt and normalize the representation of sound to improve coding efficiency (Barlow, 1961; Dean et al., 2005) and to sensitize the system to changes or new events in the acoustic environment (Stilp et al., 2010). Enhancement could therefore play an important role in everyday auditory perception. In addition, the size of the effect can be large, leading to effective amplification of the target of between 5 dB and 25 dB, depending on the task (Viemeister et al., 2013; Feng and Oxenham, 2015). Given the potential importance of auditory enhancement, it is surprising that relatively little is known about its neural origins. An earlier study in the auditory nerve of the guinea pig (Palmer et al., 1995) found that auditory-nerve fibers adapted to the precursor stimulus, so that the response to the masker was reduced more than the response to the target, leading to a relative enhancement of the target response. However, no evidence for an absolute enhancement of the target response was found, as would be needed to explain the perceptual phenomena described above (Viemeister and Bacon, 1982; Wang and Oxenham, 2016). Some enhancement has been reported in the cochlear nucleus, but it was limited to the onset of the target tone (Scutt and Palmer,

1998). In contrast, more robust enhancement effects have been found in the responses of single units within the inferior colliculus (IC) of awake passive marmoset monkeys (Nelson and Young, 2010). Taken together, the results suggest hierarchical processing, with enhancement only emerging at the level of the IC. However, such an interpretation must be tempered by the fact that the studies were carried out in different species, only the study of Nelson and Young (2010) was attempted in an awake preparation, and none of the studies included behavior. Evidence from human studies remains limited. Beim et al. (2015) found no evidence for enhancement in the cochlea, using otoacoustic emissions (OAEs), and Carcagno et al. (2014) found no evidence for enhancement in the 80-Hz auditory steady-state responses, which are thought to be primarily subcortical in origin (Herdman et al., 2002; Bidelman, 2018). Thus, it remains the case that no neural correlates of enhancement in humans have yet been reported.

Here we employed EEG to probe the neural correlates of enhancement at not only the sub-cortical but also the cortical level which has not been investigated in any previous studies. We used a stimulus design described by Feng and Oxenham (2015), which yields up to 24 dB of auditory enhancement, as measured behaviorally, potentially increasing the likelihood of observing neural correlates of the effect. In addition we used a frequency tagging paradigm that enables us to analyze the neural responses to the target and masker components separately. The target and masker components were tagged with a combination of low (around 40 Hz) and high (around 100 Hz) amplitude modulations (AMs), selected to investigate primarily cortical and sub-cortical responses, respectively. The responses to the target and masker were estimated by measuring the magnitude of auditory steady state responses (ASSRs) at these specific tagging frequencies. We observed increased responses to the target tone at both subcortical and cortical levels. The magnitude of the enhancement was larger at the cortical level, consistent with the concept of hierarchical processing, whereby enhancement is progressively increased throughout the early stages of auditory processing. Consistent with psychophysical findings, no changes in the response to the masker components were observed in the presence of a precursor.

METHODS

Participants

Ten participants (six female and four male) took part in Experiment 1 and sixteen participants (nine female and seven male) took part in Experiment 2. The participants were between 18 and 34 years old, had normal hearing, as defined by audiometric pure-tone thresholds better than 20 dB hearing level (HL) in both ears at octave frequencies from 250 to 8,000 Hz, and had no reported history of hearing or neurological disorders. All participants provided written informed consent and were compensated for their time. All protocols were approved by the University of Minnesota Institutional Review Board.

Experiment 1: Behavioral measures of enhancement

Perceptual thresholds for a target tone were measured in conditions that did and did not include a precursor stimulus, in order to confirm and quantify the amount of behavioral enhancement using the same stimuli that were then employed in the EEG recordings.

Stimuli. In the simultaneous masker condition with no precursor (MSK), each trial contained an inharmonic complex tone with five equal-amplitude components spaced apart from each other by 5/11 octaves, followed by a pure-tone probe. The target tone was the 3rd component within the complex tone. The frequency of the probe tone was either the same as the frequency of the target tone, or was geometrically centered between the frequencies of the target tone and of one of its adjacent neighbors with equal *a priori* probability. From trial to trial, the frequencies of the entire inharmonic complex were randomly roved within a one-octave frequency range (with uniform distribution on a logarithmic scale). This roving led to the frequency of the target tone being anywhere between 1 kHz and 2 kHz on any given trial. The inharmonic complex and probe tone were each 437.52 ms long, including 10-ms raised-cosine onset and offset ramps, separated by a 100-ms silent gap. The level of each masker component was 45 dB

sound pressure level (SPL). In the enhanced condition (ENH), a precursor was presented before the inharmonic complex. The four precursor frequencies matched those of the masker in each trial (i.e., no component at the target frequency). The duration of the precursor was also 437.52 ms, including 10-ms raised-cosine onset and offset ramps. The delay between the precursor offset and inharmonic complex onset was 10 ms. In the conditions with amplitude-modulated (AM) tones, the four masker components in the masker-plus-target complex were amplitude modulated with the sum of two sinusoidal waveforms at 34.28 and 91.42 Hz, each presented at a modulation depth of 25%. The amplitude modulation for the target component was the sum of two other sinusoidal waveforms at 43.43 and 98.28 Hz, each modulated at a depth of 50% (Fig. 1a). The probe tone was modulated the same way as the target. The 437.52-ms duration of inharmonic complex and probe tone ensured an integer number of cycles of all the modulation frequencies, such that the starting and ending phases were both at zero and consistent for all modulators. The precursor components were not modulated in the ENH condition (Fig. 1b).

Procedure. Participants were individually seated in a double-walled sound-attenuating booth. The stimuli were generated digitally using the AFC software package (Ewert, 2013) under Matlab (Mathworks, Natick, MA) at a 48-kHz sampling rate, delivered through an L22 soundcard (LynxStudio, Costa Mesa, CA) with a 24-bit resolution, and presented monaurally to the right ear via HD650 headphones (Sennheiser, Old Lyme, CT). The task was a present/absent task where the listeners were asked to report whether or not the probe tone was present in the target-plus-masker complex. The two alternatives (probe tone present or absent) were presented with equal *a priori* probability. The level of the target tone was initially set to 65 dB SPL (i.e., 20 dB higher than the individual masker components) and was varied adaptively following a two-down one-up rule that tracks the 70.7% correct point on the psychometric function (Levitt, 1971). Feedback was provided after each trial. The level of the probe tone was always the same as that of the target tone. Initially the level of the target was varied in steps of 5 dB. After two reversals in the direction of the adaptive tracking procedure, the step size was reduced to 2 dB. The run was terminated after eight reversals and the threshold was computed as the average target level at the last

six reversal points of the tracking procedure. There were four conditions in total, including a simultaneous masker, or no precursor, condition (MSK) and an enhanced condition (ENH) either with pure tones or AM tones. Each condition was tested once for each participant. Each participant either started with the pure tones or the AM tones, with the order counterbalanced between participants. The MSK and ENH conditions were presented in a different random order for each participant and tone type (pure or AM). Threshold was defined in terms of the target-to masker ratio (TMR), or the level of the target relative to the level per component in the remainder of the inharmonic complex.

Screening and training. Before the main experiment, participants were required to pass two pitch-discrimination training and screening sessions. In the first session, the participants were presented with two consecutive pure tones, each 437.52 ms in duration, separated from each other by a silent gap of 100 ms. The two tones were either the same or differed in pitch by the same amount as the target and probe tones in the main experiment. Participants were asked whether the two tones had the same or different pitch. In the second session, both tones were amplitude modulated with a sum of two sinusoids at 43.43 and 98.28 Hz with a 50% modulation depth for either frequency. The tones in session 1 and the carriers of the AM tones in session 2 were roved in frequency from trial to trial in the same way as in the main experiment. All participants had to obtain at least 80% correct in both sessions to pass. All 10 participants passed the screening.

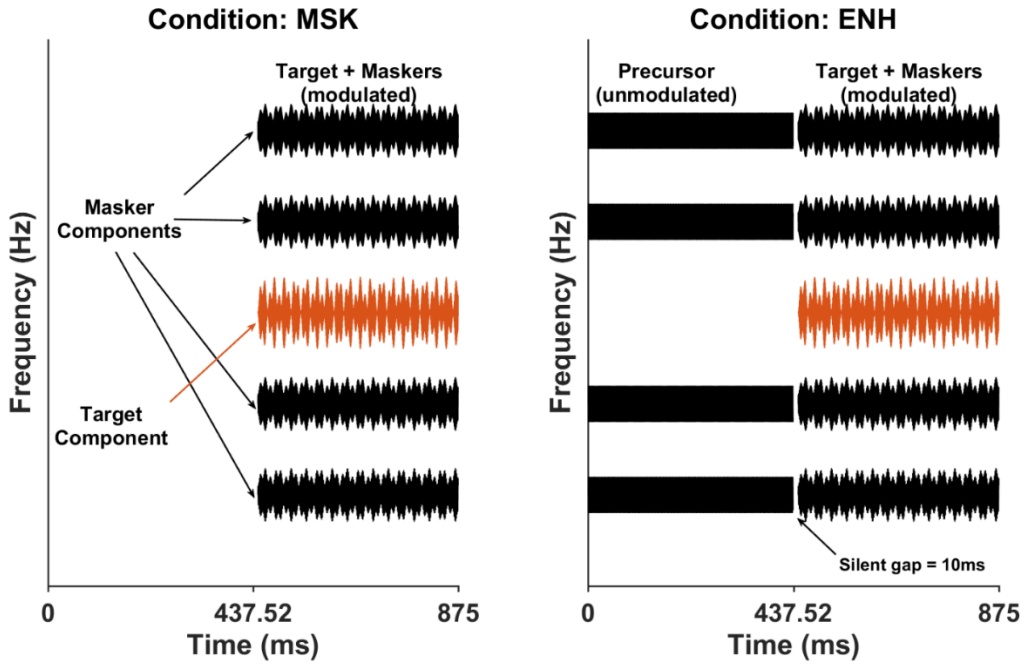
Experiment 2: EEG measures of enhancement

In this experiment, we recorded the auditory steady-state responses to estimate the population neural responses to the masker tones and target tone separately by tagging them with different signature amplitude modulation (AM) frequencies.

Stimuli. The stimuli used for the EEG experiment were the same AM tones used in Experiment 1, but without the probe tone. All four masker components in the masker-plus-target complex were amplitude

144 modulated with the sum of two sinusoids of 34.28 and 91.42 Hz with 25% modulation depth for each
145 frequency. The target component was modulated with the sum of two other sinusoids of 43.43 and 98.28
146 Hz with 50% modulation depth for each frequency. The precursor components in the ENH condition were
147 not modulated. The duration of the precursor and masker were both 437.52 ms, including 10-ms raised
148 cosine onset and offset ramps. Both conditions, with precursor (ENH) and without precursor (MSK), were
149 tested at three target-to-masker ratios (TMRs) of 0, -5 and -10 dB, resulting in a total of six conditions. A
150 total of 1000 trials were run in each condition for each participant, and the frequencies of the entire
151 inharmonic complex were randomly roved on each presentation in the same way as in Experiment 1. Half
152 of the trials were presented in the inverted starting polarity to allow for the cancellation of any stimulus-
153 related artifacts after recording (Picton et al., 1974; Skoe and Kraus, 2010).

154



155

156 ***Figure 1:*** *Schematic diagram of the stimuli with amplitude-modulated tones used in Experiments 1 and 2.*
157 *For the behavioral measurements (experiment 1), the target-plus-masker mixture was followed by a*
158 *probe tone.*

Procedure. Each participant took part in one experimental session of 2.5 hours, including behavioral measurements, setup, and EEG data collection. Participants were seated in a double-walled, electrically shielded, sound-attenuating booth. Each session started with a short behavioral test, with the same unmodulated stimuli used in Experiment 1, with one run of each of the two conditions, MSK and ENH. During the EEG data acquisition, participants were fitted with a cap (Easy Cap; Falk Minow Services) containing 64 silver/silver-chloride scalp electrodes. Two additional reference electrodes, one placed on each mastoid, and two ocular electrodes were used. The impedance of all electrodes was monitored and maintained below 10 kΩ. The EEG data were recorded at a sampling rate of 4096 Hz using a 64-channel BioSemi system. The sounds were presented via ER-1 insert phones (Etymotic Research, Elk Grove, IL), and participants watched a silent movie with subtitles during data acquisition. The six conditions (MSK and ENH conditions presented at three TMRs) were played in a different random order for each participant.

The EEG pre-processing and averaging was done using the EEGLAB toolbox (Delorme and Makeig, 2004). The raw waveforms were down-sampled to 1024 Hz, re-referenced to the average of the two mastoids, and bandpass filtered from 1 to 100 Hz using a zero phase-shift filter. For each condition, the continuous EEG time series was divided into epochs. For the MSK condition, the epoch extended from 100 ms before stimulus onset to 700 ms post stimulus onset. For the ENH condition, the epoch extended from 100 ms pre-stimulus onset to 1100 ms post stimulus onset since the stimulus was longer with the presence of the precursor. The EEG epoched signal was then baseline corrected relative to the 100-ms pre-stimulus baseline. Independent Component Analysis (ICA) was used to remove artefacts related to eye movements and blinks (Jung et al., 2000).

Further analysis was done in MATLAB. The discrete Fourier transforms (DFTs) of the processed EEG signals were applied to the time-domain waveforms from individual trials and the phases at each frequency were extracted. For each electrode in each condition for each participant, the phase locking value (PLV) to the envelope was computed by averaging the phases of the individual trials' responses at

each frequency from 400 random samples (drawn with replacement) (Zhu et al., 2013). The average phases were calculated for the 200 positive polarity trials (POS_i) and 200 negative polarity trials (NEG_i) separately beforehand (Eq. 1).

$$P_i(f) = \frac{1}{400}\left|\left(\sum_{n\in \mathrm{POS}_i} e^{i\emptyset_n(f)} + \sum_{m\in \mathrm{NEG}_i} e^{i\emptyset_m(f)}\right)\right| \quad \text{Eq. 1}$$

The same procedure was repeated 100 times independently to estimate the distribution of the PLVs and the mean was calculated as the observed PLVs for one electrode in one condition (Eq. 2).

$$P_{ENV}(f) = \frac{1}{100}\sum_{i=1}^{100} P_i(f) \quad \text{Eq.2}$$

To evaluate the statistical significance of PLV values, bootstrapping was used to estimate the noise floor. A null model was tested by generating one random distribution of PLVs by repeating the procedure for PLV calculation described above 1000 times except that the phase in Eq. 1 was set to be random (uniformly distributed from 0 to 2π, independently selected for each trial and repetition, *i*). The calculated mean PLV distributions from the experimental data can be compared to this random distribution (Bonferroni corrected for multiple comparisons): a PLV was significant if the estimate of the probability of observing it by chance was less than $p < 0.05/M$ ($M = 4$). In other words, a PLV was significant if larger than 0.075 (noise floor) in our study. For all these analyses, a subset of 28 electrodes, equally distributed across both hemispheres, was chosen for analysis. The PLVs were averaged across the subset of electrodes for each condition of each participant.

RESULTS

Behavioral thresholds: The mean behavioral thresholds for Experiments 1 and 2 are shown in Fig. 2. The mean behavioral threshold across all participants in Experiment 2 was 13.4 dB TMR in the MSK condition and -11.1 dB in the ENH condition. These results validated the TMR range (-10 dB to 0 dB TMR) chosen for the physiological measurements, showing that the selected TMRs included those for

which clear behavioral enhancement was observed. The average amount of enhancement calculated as the difference in thresholds with and without the precursors (MSK - ENH) was 24.5 dB. This value is comparable to the average enhancement reported by Feng and Oxenham (2015), showing that the modification of the current stimuli (fewer maskers and smaller roving range) did not noticeably affect the amount of enhancement. In Experiment 1, the average enhancement was ~20 dB for the modulated and unmodulated stimuli. A one-way repeated-measures analysis of variance (ANOVA) on the amount of enhancement with stimulus type (pure tones vs. modulated tones) indicated no significant effect of stimulus type ($F_{1,18}$ = 0.071; p = 0.8), suggesting that the additional amplitude modulations used for the EEG experiment did not affect the amount of enhancement.

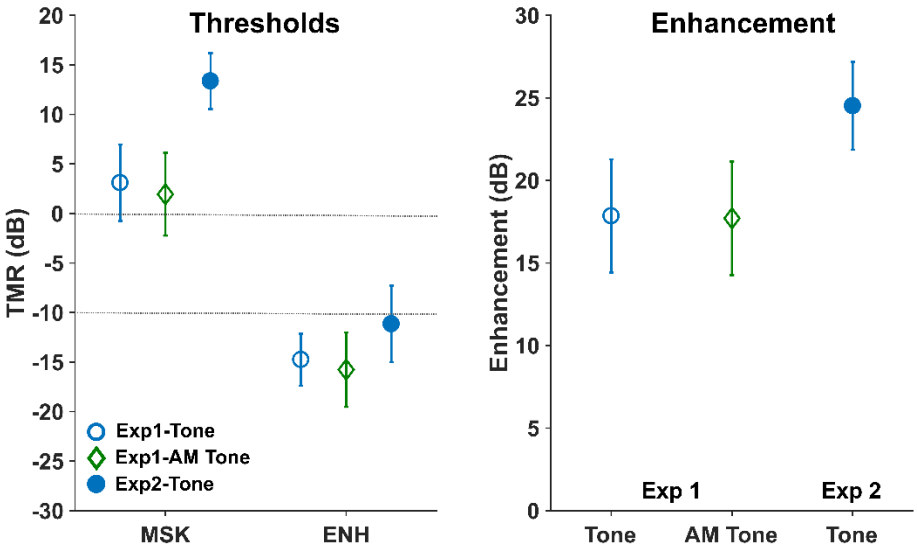


Figure 2: *Average behavioral data from Experiments 1 and 2. The panel on the left shows the raw thresholds for the MSK and ENH conditions for both Experiments 1 and 2. The region between the dashed lines shows the TMR range used for the stimuli in the EEG experiment. The panel on the right shows the average enhancement, calculated as the difference in thresholds between MSK and ENH conditions, for both Experiments 1 and 2.*

EEG data: Examples of PLVs for two representative participants are shown in Fig. 3. PLVs below the noise floor (dashed line) are not significantly greater than expected by chance. There are four distinct peaks at the four tagging frequencies for maskers and target respectively. Since the ASSR in response to AM around 40 Hz is thought to be generated in the auditory cortex (Hari et al., 1989; Herdman et al., 2002), the two peaks near 34 Hz and 43 Hz reflect cortical responses to the maskers and target respectively. Since the ASSR in response to AM around 100 Hz and above is thought to be produced primarily by subcortical generators (Bidelman, 2018), the two peaks near 91 Hz and 98 Hz are likely to reflect subcortical contributions. As shown in Fig. 3, both cortical and subcortical responses to the target component appear larger in the ENH condition than in the MSK condition.

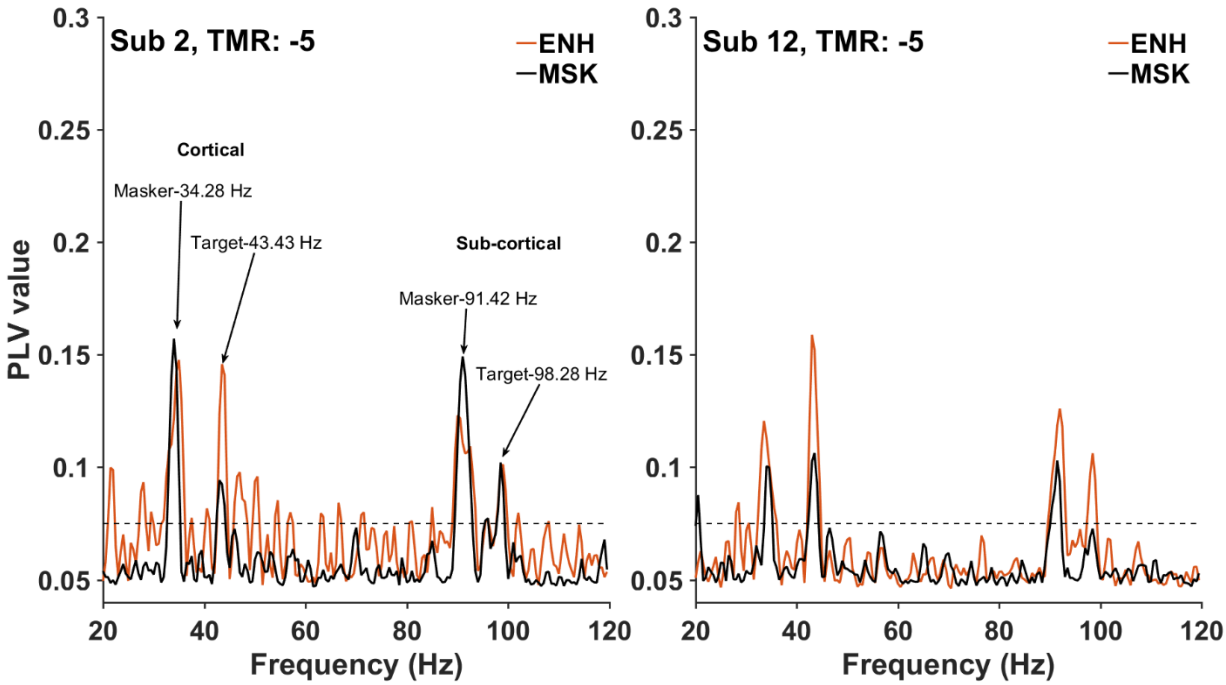


Figure 3: *Examples of PLVs from 2 subjects plotted as a function of frequency measured at TMR -5 in ENH and MSK conditions respectively. Each curve was the averaged PLVs across 28 electrodes. The arrows pointed to the four distinct peaks at all four tagging frequencies used for the maskers and target. The orange curves indicate the response in the enhanced condition (ENH) while the black curves represent the response in the simultaneous condition (MSK). Note that the amplitude of the PLV for the*

masker components remain similar in both conditions (MSK and ENH) whereas the target component amplitudes are enhanced in the ENH conditions. The dashed line indicates the noise floor.

The average responses over 17 participants are plotted in Fig. 4 as a function of the TMR. As the TMR increased, the response to the target embedded in the simultaneous masker (MSK condition) tended to increase (dashed orange lines). When a precursor was present (ENH condition), the responses to the target were enhanced compared to the MSK condition with no precursor (solid orange lines). The enhancement in the neural responses to the target appeared more pronounced at the cortical level (Fig. 4, left panel). In contrast, the responses to the masker did not vary with TMR and did not appear to be affected by the presence of the precursor (black lines). The amount of enhancement was calculated as the difference in PLV between the ENH and MSK conditions, shown in Fig. 5. Separate two-way within-subjects (repeated-measures) ANOVAs on the amount of enhancement were performed for both cortical and subcortical PLVs with factors of TMR (0,-5,-10 dB) and stimulus component (target or masker). For the cortical responses, there was a significant main effect of stimulus component ($F_{1,16} = 42.4$, $p < 0.001$). There was no significant effect of TMR ($F_{2,32} = 0.38$, $p = 0.687$) and no significant interaction between the two factors ($F_{2,32} = 0.12$, $p = 0.89$). Similarly, for the subcortical responses, there was a significant main effect of stimulus component ($F_{1,16} = 7.24$, $p = 0.016$) but no significant effect of TMR ($F_{2,32} = 0.66$, $p = 0.52$) and no interaction ($F_{2,32} = 1.58$, $p = 0.23$).

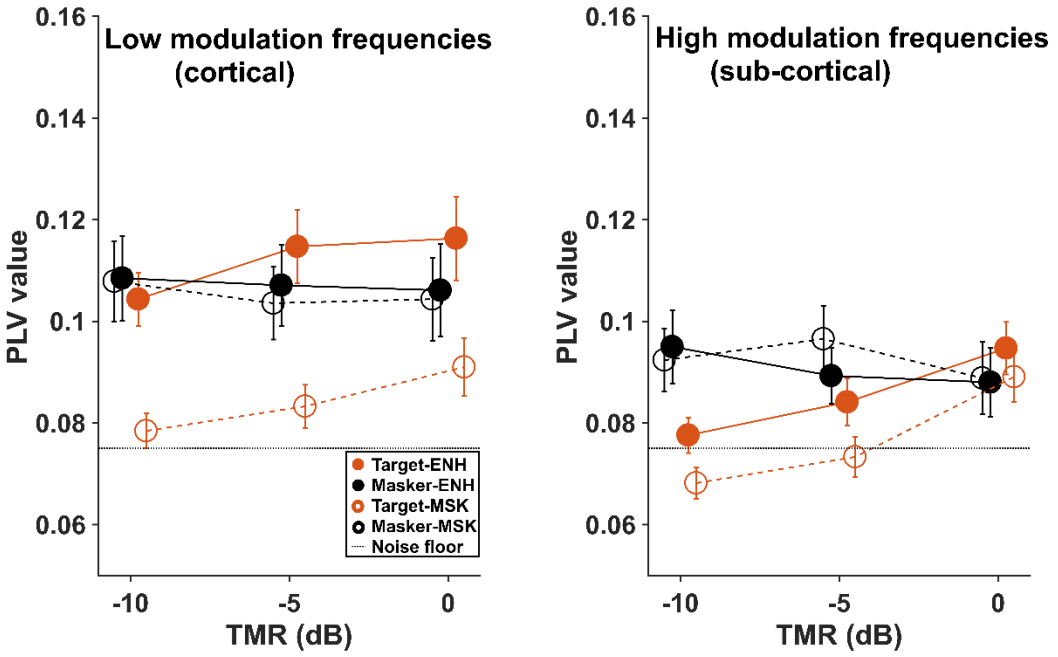


Figure 4: *Average PLVs across all participants for both low (left) and high (right) tagging frequencies for all components across TMRs. The error bars represent the standard error of the mean.*

Since the ANOVAs showed that enhancement does not depend on TMRs, we averaged the enhancement across all three TMRs for the target and masker respectively for further analysis. A one-sampled *t*-test was applied for the averaged enhancement at both the cortical and subcortical frequencies. For both cortical and subcortical frequencies, the enhancement for the target component was significantly different from zero (Cortical: $t_{16} = 7.1$, $p < 0.001$; Subcortical: $t_{16} = 2.38$, $p = 0.03$). In contrast, the enhancement of the masker components were not significantly different from zero (Cortical: $t_{16} = 0.81$, $p = 0.43$; Subcortical: $t_{16} = 0.57$, $p = 0.58$). In order to investigate whether the enhancement in target responses differs at two auditory process stages, we conducted a two-way repeated-measures ANOVA on the enhancement in target responses with tagging frequency (cortical or subcortical) and TMR as within-subjects factors. There was a significant main effect of tagging frequency [$F(1, 16) = 19.82$, $p < 0.001$]. However, neither the effect of TMR [$F(2,32) = 0.79$, $p = 0.461$] nor the interaction between the two

[$F(2,32) = 0.15$, $p = 0.86$] was significant, confirming that the amount of enhancement was smaller overall at the sub-cortical level, but that the independence with TMR was the same at both levels.

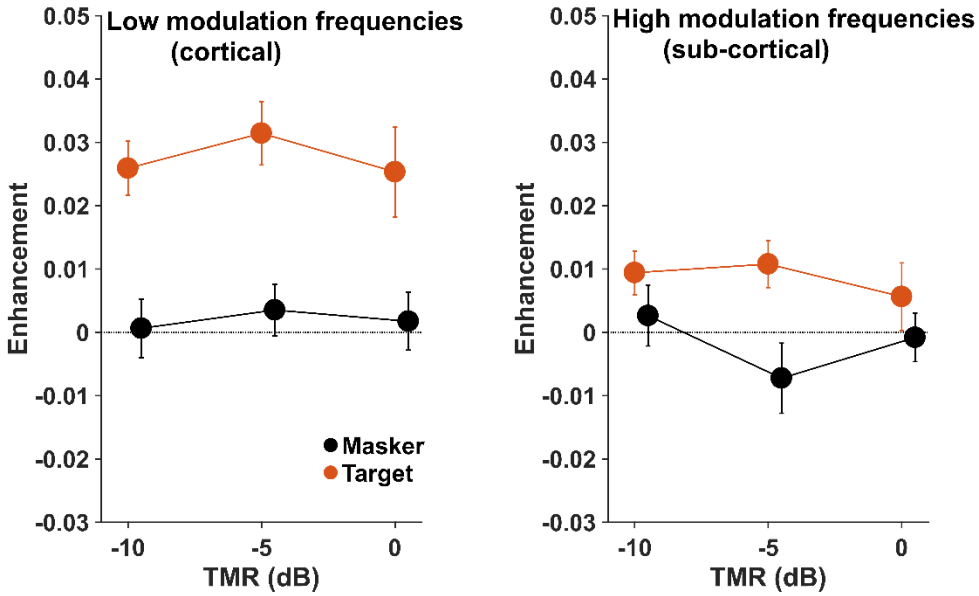


Figure 5: *Average enhancement seen as a difference in PLVs between the ENH and MSK conditions for the target and masker components across all TMRs for both low (left) and high (right) tagging frequencies. The error bars represent the standard error of the mean.*

DISCUSSION

The current study provides evidence for neural correlates of auditory enhancement in humans. These neural correlates were observed at both sub-cortical and cortical levels in humans using ASSRs with a combination of both fast (~100 Hz) and slow (~40 Hz) amplitude modulations. In line with behavioral data from earlier studies (Viemeister and Bacon, 1982; Wang and Oxenham, 2016), enhancement of the response to the target was observed together with no change in the responses to the masker. The stronger enhancement of the target with the 40-Hz ASSR than with the 100-Hz ASSR suggest stronger correlates of enhancement at the cortical than at the subcortical levels.

286 **Neural gain and behavioral thresholds**

287 The enhanced neural responses to the target in the presence of the precursor in our current study could
288 correlate with the perceptual "pop-out" of the target measured psychophysically. However, it is still not
289 clear how the neural representations are decoded or read out quantitatively to determine the perceptual
290 thresholds. For instance, in the MSK condition, the average behavioral threshold is over 10 dB TMR (Fig.
291 2), even though there is robust neural representation of the target at 0 dB TMR or lower (40-Hz ASSR).
292 Assuming that the behavioral threshold to hear out the target tone from the masker for pitch comparison
293 requires the neural responses to the target to exceed a certain threshold, the behavioral enhancement
294 should be reflected in the difference of TMRs which yield the same PLVs in the MSK and ENH
295 conditions (orange lines in Fig. 4). For instance, the average PLV of the cortical response (left panel in
296 Fig. 4) at -10 TMR in the ENH condition is equivalent to the PLV in the MSK condition with TMR $\geq$ 0
297 dB. In this case, the cortical responses would predict at least 10 dB enhancement behaviorally. In the
298 meantime, the subcortical responses (right panel in Fig. 4) would predict an effect size of approximately 5
299 dB. In this case, the predictions from the cortical responses align more closely with the 20 dB or more of
300 enhancement measured behaviorally in our current study.

301 A previous study by Carcagno et al. (2014) did not find evidence of enhancement in the 80-Hz
302 ASSRs. There are a few possible reasons to explain the apparent discrepancy between their findings and
303 ours. In their paradigm, the expected behavioral enhancement was only about 5 dB, whereas our paradigm
304 yielded 20-25 dB enhancement. Part of the difference in behavioral outcomes may be due to our use of
305 frequency roving from trial to trial, which reduces the possibility of contamination via longer-term
306 adaptation effects between trials (Feng and Oxenham, 2015). Since the change in the phase-locked
307 responses at the subcortical level might not be sufficient to account for the behavioral threshold, as
308 suggested in our current study, it is likely to be more difficult to detect the neural changes for a smaller
309 behavioral effect size. In addition, the previous study only tested one target level, which was equivalent to
310 0 dB TMR in our study. Although enhancement does not depend on TMR statistically, we did notice that

the enhancement in the subcortical response tended to be smallest at 0 dB TMR. The neural mechanisms for enhancement might operate primarily when the target level is lower than the masker level, since the inhibition effects are strongest in single neurons of central auditory system when the masker level is higher than the target level (Suga and Tsuzuki, 1985; Ehret and Merzenich, 1988; Lu and Jen, 2002).

Since the ASSRs are sensitive to loudness growth (Menard et al., 2008), the enhanced target responses in the presence of the precursor may reflect the increase in the neural gain of the target intensity. Such intensity changes might be related to the perceived partial loudness increase of the target equivalent to a 10 dB intensity change (Wang and Oxenham, 2016), as well as the increased effective level (4- 5 dB) of the target monaurally, such that a lateralized percept is produced when combining the target tone with a contralateral tone at the same frequency and phase (Byrne et al., 2011). Another study estimated that the level of the target in the MSK condition would need to be raised by 23 dB to equal the salience of the target in the ENH condition (Byrne et al., 2013). It is possible that the cortical responses also reflect (or are responsible for) the increase in saliency of the target in the ENH condition when the target is introduced within the maskers as the new event or object, which activates the bottom-up control of attention deployment (Itti and Koch, 2001; Kayser et al., 2005).

Possible neural mechanisms of auditory enhancement

One possible neural mechanism underlying auditory enhancement is the adaptation of inhibition (Viemeister and Bacon, 1982). In the central auditory system, starting from the cochlear nucleus, across-frequency processing starts to emerge, where neurons selective to one center frequency can be laterally suppressed or inhibited by neighboring frequencies (Aitkin, 1986; Rhode and Greenberg, 1994). When a complex tone is presented, the neurons that respond to each component also mutually inhibit each other. Since the frequency specific responses are known to adapt over time (Ulanovsky et al., 2003; Malmierca et al., 2009), it is possible that this form of inhibition adapts in a similar way. In the ENH condition, the

inhibition to the target response from the masker may be adapted by the precursor, resulting in an increase in the target response. Previous studies on single neurons have found evidence for this mechanism by showing increased neural firings in the cochlear nucleus (Scutt and Palmer, 1998) and inferior colliculus (Nelson and Young, 2010). In the current study, we also observed enhancement effects in phase-locked responses of neural populations in both the subcortical and auditory cortical responses of human subjects. Interestingly, our results also show that the masker responses remain unchanged in the presence of the precursor (Fig. 4). This lack of adaptation in the masker responses, which is consistent with findings of Carcagno et al. (2014), could also be explained by the adaptation of inhibition mechanism: since neurons responding to the four maskers are laterally inhibited by each other, the responses to the maskers decrease due to adaptation from the precursor, but this adaptation may be counteracted by the adaptation of the lateral inhibition, leading to no net change in responses.

Inherited or emergent?

In the current study, the enhanced neural responses to the target were reflected in both 40-Hz and 100-Hz ASSRs. However, the effects are larger in the 40-Hz ASSRs. If these responses are interpreted as emerging from the sub-cortical and cortical regions respectively, the results support the earlier indirect indications from animal studies that enhancement accumulates along the ascending auditory pathways. Although the interpretation of the 40-Hz ASSRs as reflecting neural activity in the auditory cortex is widely accepted, the origin of the 100-Hz ASSRs is still the topic of some debate. One recent study showed that cortical contributions dominate FFRs to the voice pitch (F0 = 100 Hz) of speech in MEG (Coffey et al., 2016) and this conclusion was supported by another study which showed a correlation between the strength of FFRs to F0 in EEG and BOLD signal in the right posterior auditory cortex of fMRI (Coffey et al., 2017). However, there is a natural bias of MEG to superficial brain tissue (Hillebrand and Barnes, 2002) and the relation between the BOLD signal in fMRI and underlying neural activity still remains an open question (Ekstrom, 2010). In addition, a more recent study by Bidelman (2018) showed

that subcortical structures (auditory nerves and brainstem) make the largest contribution to FFRs recorded in EEG and primary auditory cortex showed little FFR energy above 100 Hz.

It is difficult to validate the neural locus of enhancement from surface EEG measurements. It is possible that the enhanced target responses emerge in IC, as suggested by the increased amplitude of 100-Hz ASSRs in our study and enhanced firing rate of single neurons (Nelson and Young, 2010). The enhancement in neural responses in IC could be further amplified by additional cortical processes to account for the threshold changes in psychophysical measurements. Alternatively, enhancement could have a cortical origin and the neural enhancement observed in the brainstem could come from descending (efferent) corticofugal projections (Winer, 2005). Electrical stimulation of cortical neurons could result in the augmented responses of neurons with matched best frequencies, but it takes several minutes to develop (Ma and Suga, 2001). The actual sound stimulation might activate this modulatory system more quickly. Although it is difficult to fully address the question of the origin of auditory enhancement in human studies, future studies using animal models may shed some light on this important issue by examining the neural responses in the brainstem when corticofugal neurons are deactivated by cooling, optogenetic silencing, or other pharmaceutical manipulations.

To summarize, our study showed that the double-modulation tagging method can be used to probe the neural responses of individual components in a complex tone mixture at both subcortical and cortical levels simultaneously. Our results also revealed the first clear neural correlates of the important perceptual phenomenon of auditory enhancement. The fact that the enhancement was more pronounced in cortical responses than subcortical responses in our results suggests that enhancement may emerge in subcortical structures and may accumulate along the auditory pathway. The mechanisms of this accumulation, whether it is achieved via feedforward or feedback mechanisms, awaits further study.

ACKNOWLEDGEMENTS

This study was supported by NIH grant R01 DC012262. We thank Patria Leach for helping data collection.

REFERENCES

Aitkin L (1986) Discharge characteristics of units in the auditory midbrain. In: The Auditory Midbrain: Structure and Function in the Central Auditory Pathway, pp 101-128. Clifton, N.J.: Humana Press.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory Communication (Rosenblith W, ed), pp 217-234. Cambridge, MA: MIT Press.

Beim JA, Elliott M, Oxenham AJ, Wojtczak M (2015) Stimulus frequency otoacoustic emissions provide no evidence for the role of efferents in the enhancement effect. J Assoc Res Otolaryngol 16:613-629.

Bidelman GM (2018) Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. NeuroImage 175:56-69.

Byrne AJ, Stellmack MA, Viemeister NF (2011) The enhancement effect: evidence for adaptation of inhibition using a binaural centering task. J Acoust Soc Am 129:2088-2094.

Byrne AJ, Stellmack MA, Viemeister NF (2013) The salience of enhanced components within inharmonic complexes. J Acoust Soc Am 134:2631-2634.

Carcagno S, Plack CJ, Portron A, Semal C, Demany L (2014) The auditory enhancement effect is not reflected in the 80-Hz auditory steady-state response. J Assoc Res Otolaryngol 15:621-630.

Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. Nat Commun 7:11070.

Coffey EBJ, Musacchia G, Zatorre RJ (2017) Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI Evidence. J Neurosci 37:830-838.

Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 8:1684-1689.

409 Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG
410 dynamics including independent component analysis. J Neurosci Methods 134:9-21.
411 Demany L, Carcagno S, Semal C (2013) The perceptual enhancement of tones by frequency shifts. Hear
412 Res 298:10-16.
413 Ehret G, Merzenich MM (1988) Complex sound analysis (frequency resolution, filtering and spectral
414 integration) by single units of the inferior colliculus of the cat. Brain Res 472:139-163.
415 Ekstrom A (2010) How and when the fMRI BOLD signal relates to underlying neural activity: the danger
416 in dissociation. Brain Res Rev 62:233-244.
417 Erviti M, Semal C, Demany L (2011) Enhancing a tone by shifting its frequency or intensity. J Acoust
418 Soc Am 129:3837-3845.
419 Ewert SD (2013) A modular framework for running psychoacoustic experiments and computational
420 perception models. In: International Conference on Acoustics AIA-DAGA, pp 1326-1329.
421 Merano, Italy.
422 Feng L, Oxenham AJ (2015) New perspectives on the measurement and time course of auditory
423 enhancement. J Exp Psychol Hum Percept Perform 41:1696-1708.
424 Hari R, Hamalainen M, Joutsiniemi SL (1989) Neuromagnetic steady-state responses to auditory stimuli.
425 J Acoust Soc Am 86:1033-1039.
426 Hartmann WM, Goupell MJ (2006) Enhancing and unmasking the harmonics of a complex tone. J Acoust
427 Soc Am 120:2142-2157.
428 Herdman AT, Lins O, Van Roon P, Stapells DR, Scherg M, Picton TW (2002) Intracerebral sources of
429 human auditory steady-state responses. Brain Topogr 15:69-86.
430 Hillebrand A, Barnes GR (2002) A quantitative assessment of the sensitivity of whole-head MEG to
431 activity in the adult human cortex. NeuroImage 16:638-650.
432 Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2:194-203.
433 Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing
434 electroencephalographic artifacts by blind source separation. Psychophysiology 37:163-178.

435 Kayser C, Petkov CI, Lippert M, Logothetis NK (2005) Mechanisms for allocating auditory attention: an
436 auditory saliency map. Curr Biol 15:1943-1947.
437 Levitt H (1971) Transformed up-down methods in psychoacoustics. J Acoust Soc Am 49:467-477.
438 Lu Y, Jen PH (2002) Interaction of excitation and inhibition in inferior collicular neurons of the big
439 brown bat, Eptesicus fuscus. Hear Res 169:140-150.
440 Ma X, Suga N (2001) Plasticity of bat's central auditory system evoked by focal electric stimulation of
441 auditory and/or somatosensory cortices. J Neurophysiol 85:1078-1087.
442 Malmierca MS, Cristaudo S, Perez-Gonzalez D, Covey E (2009) Stimulus-specific adaptation in the
443 inferior colliculus of the anesthetized rat. J Neurosci 29:5483-5493.
444 Menard M, Gallego S, Berger-Vachon C, Collet L, Thai-Van H (2008) Relationship between loudness
445 growth function and auditory steady-state response in normal-hearing subjects. Hear Res
446 235:105-113.
447 Nelson PC, Young ED (2010) Neural correlates of context-dependent perceptual enhancement in the
448 inferior colliculus. J Neurosci 30:6577-6587.
449 Palmer AR, Summerfield Q, Fantini DA (1995) Responses of auditory-nerve fibers to stimuli producing
450 psychophysical enhancement. J Acoust Soc Am 97:1786-1799.
451 Picton TW, Hillyard SA, Krausz HI, Galambos R (1974) Human auditory evoked potentials. I. Evaluation
452 of components. Electroencephalogr Clin Neurophysiol 36:179-190.
453 Rhode WS, Greenberg S (1994) Lateral suppression and inhibition in the cochlear nucleus of the cat. J
454 Neurophysiol 71:493-514.
455 Scutt MJ, Palmer AR (1998) Physiological enhancement in cochlear nucleus using single tone precursors.
456 In: Assoc. Res. Otolaryngol. Abs., p 188(A).
457 Skoe E, Kraus N (2010) Auditory brain stem response to complex sounds: a tutorial. Ear Hear 31:302-
458 324.

459 Stilp CE, Alexander JM, Kiefte M, Kluender KR (2010) Auditory color constancy: calibration to reliable
460 spectral properties across nonspeech context and targets. Attention, perception & psychophysics
461 72:470-480.
462 Suga N, Tsuzuki K (1985) Inhibition and level-tolerant frequency tuning in the auditory cortex of the
463 mustached bat. J Neurophysiol 53:1109-1145.
464 Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat
465 Neurosci 6:391-398.
466 Viemeister NF (1980) Adaptation of masking. In: Psychophysical, Physiological and Behavioural Studies
467 in Hearing (van den Brink G, Bilsen FA, eds), pp 190-198. Delft, The Netherland: Delft
468 University Press.
469 Viemeister NF, Bacon SP (1982) Forward masking by enhanced components in harmonic complexes. J
470 Acoust Soc Am 71:1502-1507.
471 Viemeister NF, Byrne AJ, Stellmack MA (2013) Spectral and level effects in auditory signal
472 enhancement. Adv Exp Med Biol 787:167-174.
473 Wang N, Oxenham AJ (2016) Effects of auditory enhancement on the loudness of masker and target
474 components. Hear Res 333:150-156.
475 Winer JA (2005) Decoding the auditory corticofugal systems. Hear Res 207:1-9.
476 Zhu L, Bharadwaj H, Xia J, Shinn-Cunningham B (2013) A comparison of spectral magnitude and phase-
477 locking value analyses of the frequency-following response to complex tones. J Acoust Soc Am
478 134:384-395.

479