Examining replicability of an otoacoustic measure of cochlear function during selective attention

Jordan A. Beim,^{a)} Andrew J. Oxenham, and Magdalena Wojtczak

Department of Psychology, N218 Elliott Hall, 75 East River Parkway, University of Minnesota, Minneapolis, Minnesota 55455, USA

(Received 23 July 2018; revised 12 October 2018; accepted 27 October 2018; published online 20 November 2018)

Attention to a target stimulus within a complex scene often results in enhanced cortical representations of the target relative to the background. It remains unclear where along the auditory pathways attentional effects can first be measured. Anatomy suggests that attentional modulation could occur through corticofugal connections extending as far as the cochlea itself. Earlier attempts to investigate the effects of attention on human cochlear processing have revealed small and inconsistent effects. In this study, stimulus-frequency otoacoustic emissions were recorded from a total of 30 human participants as they performed tasks that required sustained selective attention to auditory or visual stimuli. In the first sample of 15 participants, emission magnitudes were significantly weaker when participants attended to the visual stimuli than when they attended to the auditory stimuli, by an average of 5.4 dB. However, no such effect was found in the second sample of 15 participants. When the data were pooled across samples, the average attentional effect was significant, but small (2.48 dB), with 12 of 30 listeners showing a significant effect, based on bootstrap analysis of the individual data. The results highlight the need for considering sources of individual differences and using large sample sizes in future investigations. © 2018 Acoustical Society of America.

<https://doi.org/10.1121/1.5079311>

[BLM]

Pages: 2882–2895

I. INTRODUCTION

In auditory perception, solving the cocktail party problem (Cherry, 1953) relies on our ability to selectively attend to one target speaker among many distracting background sounds, including other speakers. Recent work has revealed that cortical representations of unattended speech are attenuated relative to attended speech (Mesgarani and Chang, 2012). These changes are robust enough to allow computational decoding of the attended speaker using cortical activity in a closed-set environment, even within a single experimental trial (e.g., Ding and Simon, 2012; Choi *et al.*, 2013; O'Sullivan *et al.*, 2015).

Although these changes in cortical activity are robust and well documented, it is unclear at what stage of auditory processing attentional modulation is first observed. From an anatomical standpoint, attentional effects could extend to the most peripheral part of the auditory nervous system, the cochlea, via the medial olivocochlear (MOC) efferent system. The MOC system is a network of neurons in the medial part of the superior olivary complex that receives input from afferent peripheral auditory neurons and corticofugal projections. MOC efferent neurons innervate the outer hair cells (OHCs) in the cochlea (for review, see Guinan, 2006). The primary functional consequence of MOC efferent activity is a reduction of the cochlear gain produced by the OHCs.

A long line of research has investigated the role of the MOC system in human selective attention, mostly using otoacoustic emissions (OAEs; Kemp, 1978) as a noninvasive

window into cochlear function (Avan and Bonfils, 1992; de Boer and Thornton, 2007; Giard *et al.*, 1994; Harkrider and Bowers, 2009; Maison *et al.*, 2001; Michie *et al.*, 1996; Puel *et al.*, 1988; Srinivasan *et al.*, 2012; Walsh *et al.*, 2015). These studies have generally revealed small and inconsistent effects of selective attention, with effects (if any) of only a fraction of a decibel. While small effects observed through OAEs can translate to larger changes in the afferent neural signal (e.g., Puria *et al.*, 1996), the inconsistent direction of these effects makes it difficult to ascribe to them any important functional role in attention and perception.

Although these small or null results may indicate no role of the MOC system in attention, it is also possible that the failure to observe large effects has been due to the suboptimal choice of stimuli and procedures in previous studies. For instance, earlier studies have often employed relatively high sound levels, which themselves could activate the MOC system even in the absence of attentional manipulations. In addition, the attention of the participants was not always well controlled, leading to uncertainty and potential individual differences related to the degree to which attention was focused on or away from the acoustic stimuli. Specifically, conditions in which listeners performed an auditory or a visual task, typically referred to as attention conditions, were often compared with conditions in which listeners were not performing a task, referred to as inattention conditions. However, there is no guarantee that listeners did not pay attention to the stimuli when no specific task was assigned to them.

In the current study, we leveraged recent improvements in the measurement techniques used to probe MOC effects,

^{a)}Electronic mail: beimx004@umn.edu

and implemented rigorous manipulations of attention, to revisit the question of attentional modulation of human cochlear responses. Stimulus-frequency otoacoustic emissions (SFOAEs) were recorded from participants while they were engaged in behavioral tasks requiring selective attention to either auditory or visual stimuli. Care was taken in the stimulus design to produce high signal-to-noise ratios (SNRs) when measuring SFOAEs and avoid stimuli that might inadvertently activate the MOC efferent system or the middle ear reflex, for reasons discussed below. Our new combination of stimulus and attentional manipulations revealed large and significant changes in human cochlear reactions in response to changes in selective attention in the first group of participants ($N = 15$). However, a subsequent attempt at replicating this experiment in a second group of participants ($N = 15$) failed to produce the same results despite using improved instrumentation, although there remained a small but significant effect when the results from both groups were combined. Careful analysis of differences in SFOAE levels and SNRs resulting from the use of different measurement systems for the two groups of participants showed that the different outcomes could not be attributed to the differences in instrumentation. The large differences between individuals, despite relatively stable within-individual measurements, reveal inter-individual variability that remains unexplained. The outcomes highlight the need for large samples to aid in examining the large inter-individual variability observed in our study and across previous studies of attentional effects on cochlear responses.

II. METHODS

A. Overview

Participants in this study were presented with a complex set of audiovisual stimuli consisting of sequences of low- and high-frequency tones, as well as a series of visually presented letters on a computer monitor. Participants were required to complete three different tasks. The same stimuli (both auditory and visual) were present in each condition so that our critical manipulation was to vary the stimulus to which attention was directed in the behavioral task. The first two tasks required participants to attend to either a low- or high-frequency sequence of tones and report at the end of each run how many of the tones in the target sequence were modulated in amplitude. Amplitude modulation (AM) detection was chosen for the behavioral task because the MOC efferent activation has been shown to affect responses to transient or fluctuating stimuli presented in noise (Guinan, 2010). A counting task was used instead of a simple detection task with immediate responses to reduce extraneous movement and noise, which could contaminate the recording of SFOAEs. The tones were embedded in a spectrally notched noise designed to activate the MOC efferent system, as previous studies that investigated attentional modulation of acoustically elicited efferent effects (Froehlich *et al.*, 1990; Froehlich *et al.*, 1993; Maison *et al.*, 2001; Michie *et al.*, 1996; Veuillet *et al.*, 1991; Walsh *et al.*, 2014a, 2014b) have reported significant attention effects more consistently than studies in which attention effects on cochlear

responses were measured without acoustic efferent activation (Picton *et al.*, 1971; Picton and Hillyard, 1974). We also believe this creates a more ecologically valid design, as real world environments, such as a noisy restaurant, could easily contain noise levels that would acoustically elicit MOC activity. Our working hypothesis was that attention modulates the efferent effect produced by the noise (an acoustic elicitor) rather than triggering efferent activation on its own (see Fig. 6 in Walsh *et al.*, 2014a, for support of this hypothesis). The third task required participants to direct their attention to the visual stimuli. Participants monitored a stream of random serially presented letters and were instructed to detect consecutive repetitions of the same letter within the series. The visual task was chosen to compare the auditory attention conditions with a non-auditory task that also required sustained attention.

To examine the effect of attention we compared the magnitudes of the SFOAEs evoked by low-frequency tones across the three attentional conditions. A low- rather than high-frequency tone was chosen to monitor attentional modulation of efferent effects on SFOAEs because larger acoustically elicited efferent effects are observed with SFOAEs evoked by tones with frequencies in the range of 500–1000 Hz than tones with higher frequencies (Lilaonitkul and Guinan, 2012). Comparisons across attentional conditions in this study provided a window into cochlear function when attention was directed toward the SFOAE-evoking tone, toward an auditory object remote in frequency, and away from the auditory stimuli entirely.

B. Participants

Initially 18 participants (7 male, 11 female) ages 20–46 yr (median: 24 yr) were recruited to participate in the experiment. Three (all male) of the 18 participants had their data discarded due to a lack of robust SFOAEs, described in the screening procedure below, leaving a total of 15 participants (4 male, 11 female) in the first sample. A power analysis based on the variability of the data from initial participants suggested that a sample size of 15 would be required to reliably detect a 5 dB change in SFOAE magnitude between attention conditions. A second sample of 24 participants (5 male, 19 female) ages 18–34 yr (median: 21 yr) was recruited based on the power analyses to assess the reproducibility of the data from the first sample. A total of 15 participants (3 male, 12 female) in the second sample passed the screening procedures. Eight of the excluded participants did not have robust SFOAEs in the region around the nominal probe-tone frequency that was used to monitor effects of selective attention on cochlear responses. One participant was excluded due to an inability to pass the behavioral-training condition, described below. All participants were tested for normal hearing (air conduction thresholds of 20 dB hearing level (HL) or better at octave frequencies between 0.25 and 8 kHz) using a calibrated Madsen Conera audiometer (GN Otometrics, Schaumburg, IL). Participants provided written informed consent before taking part in the study and were compensated for their time.

All experimental procedures were approved by the Institutional Review Board at the University of Minnesota.

C. Auditory stimuli

Auditory stimuli were used both for evoking SFOAEs and for the auditory behavioral task. Ten blocks of task-relevant stimuli were grouped to form a single run (a single row of stimuli in Fig. 1). Every run began with a baseline SFOAE measurement utilizing the suppression method (Shera and Guinan, 1999). A 10-s probe tone at the frequency selected by the screening procedure (~ 0.75 kHz) was presented at 35 dB sound pressure level (SPL) with a suppressor tone 50 Hz higher in frequency and presented at 55 dB SPL. The suppressor tone was gated on and off every 0.5 s, starting 0.5 s after the onset of the 10-s tone. The baseline SFOAE magnitude was used to ensure that an SFOAE at the probe frequency has a sufficiently high SNR to observe robust efferent effects (Goodman *et al.*, 2013). The baseline was also needed to ensure that SFOAE magnitude due to the acoustically elicited medial olivocochlear reflex (MOCR) allowed room for attentional modulation of efferent effects (i.e., the cochlear gain reduction by the MOCR did not obliterate the probe-evoked SFOAE).

The task-relevant stimuli began after a 2-s silent gap following the offset of the probe tone. After the silent gap, a 1-s cue tone was presented at either the probe-tone frequency or 4 kHz to indicate which frequency participants should attend to complete the perceptual auditory task (AM detection). A spectrally notched Gaussian white noise with 1-octave-wide

notches centered on both the probe-tone frequency and 4 kHz began 0.5 s after the onset of the cue tone and was presented for 2 s. The overall root-mean-square (rms) level of the noise was set to 55 dB SPL. The function of the noise was to activate the MOC reflex; according to our working hypothesis, attention should modulate the change in cochlear gain produced by MOCR activation. A total of four 0.3-s tone pips (two at each frequency) were presented within the duration of the noise in each block. Tone pips of the same frequency had randomly selected inter-stimulus intervals (ISIs) to ensure sustained attention by preventing listeners from attending only at the expected times. The ISIs were constrained to have a minimum duration of 100 ms and ensure that the final-tone offset occurred no later than the offset of the noise. The pips were presented at 35 dB SPL. This low level reduced the possibility of the pips activating MOC efferents. It was important to avoid efferent activation by the pips because the efferent effect on subsequent pips in a run would vary depending on the ISI and could introduce unwanted variability and, thus, contaminate effects of attentional modulation. The tone pips and noise were gated with 5-ms onset and offset cosine-squared ramps to reduce spectral splatter. The high- and low-frequency tones could overlap in their presentation time so that performance in the task would benefit from efferent inhibition of cochlear responses in the unattended frequency region. The task-relevant stimuli ended with a 2-s silent interval.

The target tone pips for the behavioral task (between one and five within each run) were amplitude modulated for

FIG. 1. A schematic illustration of the stimuli used for evoking SFOAE and for the auditory and visual behavioral tasks. Time-frequency representations of auditory stimuli are shown across multiple runs (ten total runs per attention condition). The baseline SFOAE measured by a continuous probe is illustrated in dark blue. Successive presentations of the suppressor (in black) were presented in alternating phase. Task relevant auditory stimuli, depicted in the lightly shaded region, were presented simultaneously with the visual stimuli. A repeating sequence (block) of sound follows the baseline measurement. Each block contains a cue tone and two target tone pips (dark blue), two distractor tone pips (light blue), and notched noise (dark grey). Modulated target tones that the listener counted while participating in the attend-low condition are shown in red. The inset at the bottom is a series of panels depicting the location and timing of visual stimuli on the computer screen. Each panel represents the display at one moment during visual stimulus presentation. The duration of each frame is presented on the screen and shown below the frame.

their full duration at a rate of 10 Hz with a modulation depth of -5 dB expressed in units of $20 \log(m)$, where m is the modulation index between 0 (no modulation) and 1 (100% modulation). A total of ten recordings (from ten runs) were made for each experimental condition. A schematic representation of the auditory stimuli is shown in the top section of Fig. 1. All auditory stimuli were presented to the right ear of each participant. Visual stimuli shown at the bottom of Fig. 1 were also presented, as described in detail below.

All stimuli were generated digitally with a sampling frequency of 48 kHz and a 24-bit depth using custom MATLAB software (The Mathworks, Natick, MA), converted into analog signals using a LynxTwo-B sound card (Lynx Studio Technology, Costa Mesa, CA), and delivered to participants via transducers in either an ER10C or ER10X system (Etymotic Research, Elk Grove, IL). The ER10C system was used to present and record stimuli for the first group of participants, while the ER10X system was used for the replication sample. Recordings were made using the microphones of the ER10C or ER10X systems. Each sound system provided a 20 dB gain to the microphone signal before it was digitized by the LynxTwo-B sound card and saved on the computer hard drive for offline analysis. Both sound delivery systems were calibrated with a type 4153 artificial ear with a 2 cc coupler (Bruel and Kjaer NA, Duluth, GA) to verify output sound pressure in an average ear canal.

D. Visual stimuli

The visual stimuli were presented concurrently with the auditory stimuli (i.e., beginning after the 2-s gap following the baseline SFOAE measurement as shown in the bottom row of Fig. 1) and consisted of 35 letters presented serially. The letters were drawn randomly from the first 11 letters of the English alphabet. This subset of letters was chosen because random sampling with replacement from this set produced, on average, three pairs of consecutive letter repetitions (the same number as average target tone modulations in auditory stimuli). Letters were displayed on the screen for 0.3 s with random ISIs (mean 1.7 s), but the intervals were constrained so that the total visual stimulus presentation time exactly matched the total duration of the auditory stimuli in a run. The location of each letter on the screen was randomly selected. Letters were presented at 80% contrast and subtended approximately two degrees of visual angle. The horizontal and vertical dimensions of the screen (and, hence, possible locations of the letters) subtended approximately 35 and 26.25 degrees, respectively. A schematic illustration of the visual stimuli is shown in the bottom row of Fig. 1.

E. Procedure

All procedures described below were conducted in a double-walled sound attenuating chamber (Industrial Acoustics Company, Bronx, NY) with the participant seated comfortably in a semi-reclined position in a chair with a head rest to reduce movement-related artifacts. Participants kept a computer keyboard on their lap for responses at the end of each run to minimize movement between recordings,

and were instructed to remain as still and relaxed as possible without closing their eyes or falling asleep.

Before the testing began, participants underwent a two-part screening procedure to ensure that (1) spontaneous otoacoustic emissions (SOAEs), if present, were remote from the experimental frequency range and (2) robust SFOAE measurements were obtainable within the range of the experimental stimuli. Spontaneous emissions were measured using a procedure described by Penner *et al.* (1993). To screen for robust emissions, SFOAEs were measured using the suppression method across a frequency range from 675 to 825 Hz in 15-Hz steps. In order to pass the screening, SFOAEs needed to be at least 20 dB greater than the measurement noise floor for at least one of the frequencies used, and the phase coherence of SFOAE needed to be greater than 0.95. These parameters were chosen in order to maximize the SNR of the measurements, as high SNR measurements are necessary to reveal significant MOC efferent effects (Goodman *et al.*, 2013). For any given participant, the frequency with the largest SFOAE magnitude in this screening was used as the SFOAE probe frequency in the experiment. Three participants were rejected from the total sample because they did not show SFOAEs with $\text{SNR} > 20$ dB. No participants were rejected due to SOAEs near the experimental frequency range.

After passing screening, participants were initially exposed to a less complex set of the experimental stimuli to orient them to the task. First, audio examples with tone pips of just one frequency were played to participants until they could correctly identify modulated tone pips to the experimenter. The distractor tone pips at the other frequency (with and without modulation) were then added, followed by the addition of the notched noise. After it was clear that the participants understood the behavioral task and were familiarized with all the stimuli that they were supposed to attend to or ignore in the actual experiment, the next step was to complete a set of ten training runs for each behavioral condition before beginning the experiment. Training runs utilized the same stimuli as in the actual experiment, except modulation never occurred at the distractor frequency and no MOC-reflex eliciting noise was presented. If task performance during these training runs was too poor, the acoustic stimulus was reduced in complexity (in reverse of the procedure described above) until performance was above 70% correct. This percent correct means that participants reported a correct number of AM tones on at least 7/10 runs (the actual percent of correct AM detections was likely higher). Participants typically completed the training in the first ten runs, and training lasted 30–60 min. The training and screening procedures were typically completed during one 2-h experimental session.

In a second experimental session after completing the training, participants completed runs of experimental trials under three different attentional manipulations. Participants were required to attend either to the sequence of low-frequency (hereafter AL) or high-frequency (AH) tone pips, or to attend to the sequence of visually presented letters (V) on the computer screen and perform the relevant behavioral task for the attended stimulus (i.e., counting AM tones in the

AL and AH conditions, and counting consecutive letter repetitions in the V condition). Responses were made at the end of the run to ensure as little movement as possible during SFOAE recording. Ten experimental runs were recorded for each condition, providing a total of 100 presentations of the SFOAE stimuli for analysis. Experimental runs were grouped by attention condition, and the order in which attention conditions were presented was counterbalanced across participants to attempt to control for any effects that may slowly build up over the course of the experiment. Participants completed all three attention conditions in a single session, during which the probe assembly was not removed from their ears.

F. SFOAE analysis

Before emission extraction, audio recordings were manually scanned for noise and movement-related artifacts. Recorded runs containing visible artifacts were excluded from analysis. Additional recordings were made if the magnitude of SFOAEs was less than 20 dB above the noise floor. On average 7 recordings (containing approximately 70 probe-suppressor pairs for the baseline measurement) per attention condition per participant were used in the following analyses. The numbers of artifact-rejected recordings were very similar across the different attention conditions.

To estimate the SFOAE magnitude evoked by the task-relevant stimuli, the SFOAE from the baseline measurement was extracted utilizing the suppression technique (Shera and Guinan, 1999). First, the 1-s segments of SFOAE baseline measurement stimuli consisting of repetitions of 0.5-s SFOAE probe and 0.5-s SFOAE probe plus suppressor were averaged together. Next, a heterodyne procedure was used to extract the complex-valued sound pressure in the ear canal at the stimulus frequency (Guinan *et al.*, 2003). A phasor diagram depicting the vector subtraction resulting in the extracted emission is shown in Fig. 2. The magnitude and phase of this sound pressure from an example listener is shown in Figs. 3(A) and 3(C). The suppression method of SFOAE extraction uses a vector difference between ear canal sound pressure recorded with the SFOAE probe tone alone and the sound pressure recorded with the SFOAE probe tone and suppressor to estimate the SFOAE residual. The resultant is termed the residual since the suppressor tone may not be completely effective in suppressing the SFOAE. The mean complex-valued sound pressure during a 300-ms window temporally centered in the duration of the suppressor tone was subtracted from each point of the heterodyned waveform. This analysis yields the magnitude and phase of the SFOAE residual during the first 500 ms of the waveform when no suppressor was present and the noise floor during the final 500 ms, as shown in Figs. 3(B) and 3(D). The same mean complex-valued sound pressure was also subtracted from the heterodyned average sound pressure of the low-frequency tone pips used during the behavioral task. Ten tone pips (one per behavioral stimulus block) were averaged together across blocks and runs to yield a single average tone pip. Tone pips that contained amplitude modulation (behavioral targets) were always excluded from the

FIG. 2. Phasor diagrams and time-frequency schematic illustrations depicting the emission extraction procedure from the averaged temporal waveform of recorded audio. (A) A phasor diagram depicting the extraction of baseline SFOAE as the resultant of vector subtraction of baseline pressure P_b and suppressed baseline P_s . (B) Phasor diagram depicting the extraction of SFOAE from the task relevant stimuli as the vector difference between pressure during the behavioral task P_t and P_s . (C) and (D) are time-frequency schematics of the stimuli used to extract SFOAEs depicted in (A) and (B), respectively. Shaded grey regions represent the notched noise elicitor. Dashed boxes highlight analyses windows where averaging occurs to perform the vector subtraction.

averaging procedure. If neither of the two tones within a block contained amplitude modulation, the first tone was always chosen. The level of SFOAE residual obtained from the tone pips during a 200-ms window in the temporal center of the tone-pip is shown in Fig. 3(E) (in red) with the level of SFOAE residual during the baseline measurement (in grey). Predictably, the level of the residual during the task is lower than the baseline due to MOC efferent activation by the notched noise presented with the test tones, but it is well above the noise floor [compare Figs. 3(B) and 3(E)], leaving room for attention-related changes in MOC efferent effect on SFOAE magnitude.

G. Bootstrap analysis procedure

In order to determine whether differences in emission level during the behavioral tasks across attentional conditions were significant within an individual, a bootstrapping procedure was used to estimate the variability in emission magnitudes. Artifact-free pairs of recording segments for both the baseline and task measurements were randomly sampled with replacement, such that the total number of resampled segments was always equal to the total number of original segments. The SFOAE extraction procedure detailed above was repeated on each resampling, and the mean emission magnitudes were saved. The resampling and analysis procedure was repeated 10 000 times per participant to estimate distributions of emission magnitudes across the attentional conditions.

FIG. 3. Example extracted SFOAE data from a single listener (S12) during the AL experimental condition. Dashed lines in (A)–(D) depict the onset time of the tone used to suppress the emission. Panels A and B show extracted levels of ear canal sound pressure at the stimulus frequency and SFOAE, respectively. (C) and (D) depict phase of the same signals. (E) shows the level of SFOAE during the baseline (grey) and behavioral (blue) task segments during the 200-ms window used to compute average magnitudes for each listener and condition.

III. RESULTS

A. Behavioral results

Since participants responded with counts of detected targets at the end of a run instead of at the moment of target detection, it is not possible to perform many detailed analyses on the behavioral responses. Although this limits our ability to correlate our behavioral and OAE metrics, we still provide evidence that our listeners could complete the behavioral tasks with reasonable accuracy and our two samples did not differ in their performance of the task.

1. Original group

Participants responded with the exact count of targets in 51.7% of runs across all three conditions. Because responses in single runs were aggregate counts of the number of detected modulations over ten stimulus blocks, this metric is not the same as the typical percentage of correct detections as even a single miss would cause a block of nine other successful trials to be counted as incorrect. To provide a secondary account of accuracy in the behavioral task, we calculated the participants' root-mean-square error (RMSE) from the correct number of presented targets. This measure better compares counts of responses to targets across runs. On

average, participants' RMSE was 0.84 targets across all three attention conditions. Based on this we are confident that participants were able to attend the required stimuli, considering that a total of 20 (10 at each frequency) tone pips are presented within a single run and up to 10 (5 at each frequency) could be modulated and listeners could respond with any number. A repeated-measures analysis of variance (ANOVA) revealed no significant effect of task on behavioral performance after applying a Greenhouse-Geisser correction for a violation of sphericity [$F(1.32, 18.8) = 3.81$, $p = 0.056$]. The nearly significant effect of task on performance was driven by lower performance in the AL (mean RMSE: 1.21) condition relative to AH (mean: 0.79) and AV (mean: 0.52). Because modulation depth was set to make AM detection challenging (recall that not all instances of AM at the cued frequency were detected for the stimuli without the distractor AM and the notched noise during the training), the less-than-perfect performance does not necessarily indicate slips in attention to the cued frequency in the auditory tasks.

2. Replication group

The second (replication) group of participants responded with correct target counts in 52% of runs. Their RMSE was 0.81 across all three conditions. A repeated-measures ANOVA using just the participants from the second group revealed no significant effect of task on performance [$F(1.46, 20.4) = 2.97$, $p = 0.09$]. While the effect of task on performance was not significant, there was a trend for lower performance in the AL (mean: 1.08) relative to the AH and V conditions (means: 0.67 and 0.68, respectively), as found in the original group.

3. Group comparison

A mixed repeated-measures ANOVA, including a between-subjects factor of experimental group, showed no significant effect of experimental group on performance [$F(1, 28) = 0.049$, $p = 0.827$]. There also was no significant interaction between experimental groups and task [$F(1.45, 40.5) = 0.536$, $p = 0.532$]. There was a significant effect of task on performance in the pooled results from the two groups [$F(1.45, 40.5) = 6.55$, $p = 0.008$], based on the lower performance in AL condition than in the AH and V conditions.

B. SFOAE results

1. Original group

First, a within-subjects analysis was undertaken, using the bootstrap procedure described above. The bootstrap distributions were used to estimate 95% confidence intervals around the mean emission magnitudes; selective-attention effects within a participant were considered significant if the median emission magnitude in the AH or V conditions (when attention was directed away from the SFOAE-evoking tone) fell outside the 95% confidence interval for the magnitude of emission in the AL condition (when attention was directed toward the SFOAE-evoking tone). Nine

out of 15 participants demonstrated a significant change in emission level between the AL and V conditions, all with the emission levels lower (indicating more efferent inhibition of cochlear responses to the low-frequency tone) in the V condition than in the AL condition. Seven participants had significantly different emission levels in the AH condition compared to the V condition with the AH condition having higher emission levels in 6/7 cases. The two auditory conditions (AL and AH) had significantly different emission levels in 9/15 cases with the AH condition showing greater efferent inhibition than the AL condition in 5 of those cases.

The top panels in Fig. 4 show results of the bootstrapping procedure from a representative participant with a significant effect and a representative participant with no significant effect. Each panel contains histogram distributions of SFOAE and noise floor magnitude in the AL and V conditions. Six participants did not show significant differences in emission magnitude across conditions, but the results often exhibited the same trend of lower emission levels during the V than during the AL condition. We emphasize the AL and V conditions here for plot clarity since we observed the largest differences between these two conditions. Note that in all cases shown in Fig. 4 there is an almost complete overlap of the distributions representing the noise floor, indicating that a change in the level of the noise floor cannot account for the differences between the conditions.

Next, effects of condition were evaluated across participants. The individual SFOAE magnitudes for each attention

condition, as well as the mean across participants, are shown in the top panel of Fig. 5. The patterns of results depicted in the bootstrap analysis can be seen across attention conditions. Participants for whom the bootstrap procedure found significant differences in emission between AL and V are denoted with asterisks above the bars. As shown by the right-most set of bars, the SFOAE magnitude was on average smallest for the V condition, highest in the AL condition, and intermediate in AH condition.

A Bayesian analysis procedure (Kruschke, 2013) was used to determine whether differences in emission magnitudes between attention conditions were credibly different from zero. Briefly, SFOAE magnitudes from the task stimuli were subtracted across pairs of conditions (e.g., AL – V) for each participant. These differences were modeled using a t -distribution and posterior distributions of the mean, standard deviation, and degrees of freedom of the t -distribution were simulated using custom MATLAB software (The Mathworks, Natick, MA) to perform the Markov-Chain Monte Carlo simulation via Gibbs sampling. The 95% highest density interval (HDI) of the posterior distribution of the mean difference between each pair of attention conditions was compared to zero to determine the credibility of the difference in emission magnitudes. The analysis found a credible difference in emission magnitudes between AL and V (mean = 5.3 dB) conditions as well as a credible difference between AH and V (mean = 3.84 dB) conditions, but not between AL and AH conditions. The posterior distributions of the mean difference in SFOAE magnitude for each pair of conditions and the 95% HDI are shown in the top panels of Fig. 6. A repeated-measures ANOVA also confirmed a significant effect of attention condition on emission magnitude [$F(2,28) = 9.37$, $p = 0.001$], and paired-samples t -tests confirmed the same significant pairwise differences between the V condition and both the AL [$t(14) = 4.160$, $p = 0.001$] and the AH [$t(14) = 3.199$, $p = 0.006$] conditions.

2. Replication group

The results of the SFOAE analyses for the same within- and between-subjects comparisons as described in Sec. III A are shown in the bottom halves of each of the Figs. 4–6. As shown in Figs. 4–6 (e.g., top vs bottom panel of Fig. 5), there was a difference in SFOAE levels between the original and replication group. However, the SFOAE in the replication group remained sufficiently high to detect potential attentional modulation. For instance, S2 and S10 in the original group show large differences in emission level between AL and V but also have considerably weaker emissions than S4 and S15 in the replication group. This issue is addressed further in Secs. III F and IV B 2.

The bootstrap analysis revealed significant changes in emission levels between the AL and V conditions in only 3 out of the 15 participants in the replication group. In contrast to the original group, all three participants showed small increases in emission levels in the V condition relative to the AL condition. Bootstrapping did not reveal any individually significant differences in emission levels between the AH and V nor the AL and AH conditions. The Bayesian analysis revealed that there were no credible differences in emission

FIG. 4. Results from the bootstrapping procedure used to quantify measurement variability within participants. Histogram bars depict the mean level of the noise floor NF and SFOAE in both the AL and V conditions. Top panels show distributions for representative participants from the original sample, bottom panels show distributions for participants from the replication sample. Left panels show examples of data where no significant change existed across attention conditions, while the right panels are typical of participants who displayed an individually significant difference between the AL and V conditions. Vertical dashed lines show the mean of each SFOAE distribution. Light blue horizontal bars under the AL distributions indicate a 95% confidence interval around the mean of the distribution.

FIG. 5. Individual SFOAE levels across attention conditions obtained from the standard heterodyne procedure (not the bootstrap procedure). Asterisks denote participants for whom the bootstrap procedure found significant differences in SFOAE level between the AL and V conditions. Mean values across participants with standard error are shown by the right-most set of bars.

level between any of the three attention conditions across participants. This was also confirmed by a repeated-measures ANOVA [$F(2,28) = 0.932$, $p = 0.408$].

3. Group comparison

A mixed ANOVA was used to compare the effects of the within-subjects factor of attention condition and between-subjects factor of group to examine whether attention effects were similar between the original and replication groups. The ANOVA revealed a significant main effect of condition in the combined sample [$F(2,56) = 5.492$, $p = 0.007$]. This effect was driven primarily by the difference in SFOAE magnitude between the AL and V conditions (mean AL - V = 2.48 dB). There was also a significant effect of group on SFOAE

magnitudes [$F(1,28) = 6.69$, $p = 0.010$]. Participants in the replication group had smaller SFOAE magnitudes (mean difference between the two groups: 5.46 dB). There was a significant interaction between experimental group and attention condition [$F(2,56) = 8.67$, $p = 0.001$]. The interaction confirms that the effect of attention differed between groups (i.e., there was a main effect of attention condition in the original but not the replication group).

C. Relationship between behavioral and SFOAE data

To determine whether the strength of attentional modulation of the SFOAE was related to behavioral task performance, the difference in SFOAE magnitude between each auditory condition and the visual condition (AL-V and

FIG. 6. Posterior distributions of the mean difference in SFOAE magnitude between pairs of attention conditions from the Bayesian analysis of group data. The shaded horizontal bar on the x axis denotes the 95% HDI used to assess the credibility of the differences in each pair. Vertical dashed lines at $x = 0$ help to visualize how each distribution relates to a 0-dB effect. If the bar illustrating the 95% HDI does not overlap with the dashed line, the effect can be considered credibly different than zero.

AH-V) was compared to performance in each condition, in terms of the RMS error of responses to the correct number of targets for each condition. We focus on the difference between these two pairs of conditions as they produced significant differences in SFOAE magnitude in the original experimental group.

There was no significant correlation between SFOAE AL-V magnitude difference and the behavioral performance in the AL and V tasks [$r(28) = 0.08$, $p = 0.68$ for the auditory task, and $r(28) = -0.23$, $p = 0.23$ for the visual task]. There were no significant correlations between performance in any condition and AL-V difference in SFOAE magnitude in either subgroup, although there was a trend toward better performance in the visual task being related to larger differences between SFOAEs during V vs AL attentional conditions in the original sample [$r(13) = -0.74$, $p = 0.08$]. There were also no significant correlations between SFOAE AH-V magnitude difference and the behavioral performance in the AH [$r(28) = -0.15$, $p = 0.428$] and V tasks [$r(28) = -0.26$, $p = 0.165$]. These correlations were also not significant within each subgroup.

D. Noise floor and baseline emission measurements

To ensure that changes in emission were related to differences evoked by task demands, we recorded the level of the baseline SFOAE as well as the noise floor of the measurements from a time when the suppressor tone was present. These metrics should remain constant across attention conditions and changes in either baseline SFOAE or noise floor measurements could indicate contamination of our selective attention effect by artifacts. A mixed repeated-measures ANOVA confirmed that there was no significant effect of attention condition on the baseline SFOAE magnitudes [$F(2,43.6) = 0.621$, $p = 0.50$] or the noise floor [$F(2,56) = 0.29$, $p = 0.75$] across both experimental groups. There were no interactions between the selective attention effect and experimental group on these measures, {[$F(1.56,43.55) = 2.76$, $p = 0.072$] and [$F(2,56) = 0.266$, $p = 0.767$] for baseline SFOAE magnitudes and noise floor, respectively} but there was a group difference in baseline SFOAE magnitudes [$F(1,28) = 8.57$, $p = 0.007$].

E. Changes in emission magnitude over time

Each participant completed the three experimental conditions in a randomized order. Since all conditions were run within a single session without removing the probe, slow systematic variations in SFOAE magnitude due to drifts, for example, could confound selective attention effects. To address this possibility, we investigated changes in the baseline SFOAE measurement across condition order, rather than attention condition, to test for potential slow changes in emission magnitude over time spanning all attention conditions. A repeated-measures ANOVA with a between-subjects factor of experimental group showed that measurement order (or the time elapsed since the start of data collection) did not significantly influence the noise floor, baseline SFOAE, or task SFOAE [across all three tests, $F(2,56) < 2.75$, $p > 0.07$]. There were no significant interactions between group and

order [$F(2,56) < 2.25$, $p > 0.12$]. Mean magnitudes across participants are shown in Fig. 7(A). We also examined stimulus pressure in the ear canal within each condition in the participants who showed the largest AL-V differences and found no readily apparent systematic change in ear canal sound pressure over time. Figure 7(B) shows ear canal SPL as a function of run number for a representative participant for each of the three attention conditions (S1 from the original group).

F. Group differences

Due to the large difference in the observed selective attention effect between the two groups, we compared demographic data collected from the groups during intake. The groups differed slightly in age, with the original group being slightly older than the replication group (difference in median age = 3 yr). Age was not significantly correlated with the AL-V SFOAE magnitude difference. The groups also differed slightly in racial makeup, with the original group containing a larger proportion of participants who self-identified as Asian/Asian-Pacific Islander (hereafter, Asian) on our participant intake questionnaire (11/15 in the

FIG. 7. Effects of time on emission measures. (A) shows bar plots of the average magnitude of the noise floor, baseline SFOAE, and task SFOAE, for the first, second, and third measurements, respectively. Error bars denote standard deviation. (B) plots the stimulus ear canal pressure in a representative participant (S1, original group) as a function of run number and attention condition. Numbers next to the attention condition labels in the legend denote the order in which conditions were run.

original sample vs 6/15 in the replication). When grouping all of the participants by race, those who identified as Asian did have larger differences in SFOAE between the AL and V conditions than those who identified with other racial descriptors (3.55 dB and 1.07 dB, respectively), but this difference between groups was not statistically significant [$t(28) = -1.34$, $p = 0.19$]. There were no significant differences in SNR for baseline SFOAEs between Asian and non-Asian participants [$t(28) = -0.92$, $p = 0.37$].

Although handedness has been shown to influence the strength of MOC function (Khalfa *et al.*, 1998), it was not assessed in our participants. It is possible that a large number of left-handed individuals would possess weaker MOC effects in their right ears (potentially leading to a group difference), but it seems unlikely that either group would contain a large number of left-handed participants given the relatively low incidence of left handedness in the general population (Gilbert and Wysocki, 1992).

As mentioned in Sec. II (Methods), participants in the original and replication groups were tested using different instruments (ER10C vs ER10X, Etymotic Research, Elk Grove, IL). A consequence of using different instrumentation was that the groups differed in the measured stimulus-frequency ear canal pressure (mean: original - replication = 7.49 dB). Consequently, the original group had significantly larger SFOAEs (both baseline measurement and during the task) than those in the replication group (mean: original - replication = 5.62 dB) and larger SNRs (mean: original - replication = 7.56 dB). Larger SFOAE levels and SNRs may have increased the chance of observing selective attention effects, and thus may have contributed to the different outcomes for the two groups. To examine this possibility, the magnitude of differences in SFOAE levels between the two auditory conditions and the visual condition (AL-V and AH-V) were correlated with the measured ear canal SPL at the stimulus frequency and the result is shown for the participants pooled across the two groups in Fig. 8(A). If the lower level stimuli limit the ability to see MOC effects, then the differences in SFOAE between attention conditions should be smaller at lower sound levels. There was no significant

correlation between ear canal pressure and the size of the difference in SFOAE between conditions for neither AL-V [Fig. 8(A), all dark blue symbols; $r(28) = 0.255$, $p = 0.173$] nor AH-V [all light blue symbols; $r(28) = 0.325$, $p = 0.08$]. Within each group, there was also no significant relationship between ear canal sound pressure and AL-V, but the data trended toward a negative relationship. In the original group there was no correlation between ear canal pressure and AH-V [Fig. 8(A), light blue open filled; $r(13) = -0.119$, $p = 0.672$]. There was a significant correlation between ear canal pressure and AH-V [Fig. 8(A), light blue open; $r(13) = -0.825$, $p = 0.0002$] in the replication group, but this would suggest that smaller effects were seen with increasing ear canal pressure, which is opposite to the overall trend in the data and would not explain why greater attention effects were seen in the original group.

Although the noise floor did not differ between the two groups, the difference in stimulus pressure in the ear led to a difference in SNR for the SFOAE measurements between groups, as confirmed by an ANOVA [$F(1,28) = 7.77$, $p = 0.01$]. The ANOVA results did not indicate that SNR varied across conditions [$F(2,56) = 0.17$, $p = 0.84$]. There was no interaction between SNR and the experimental group [$F(2,56) = 0.04$, $p = 0.96$]. A scatterplot of the individual SNRs for the corresponding selective attention effects are shown in Fig. 8(B). There were no significant correlations between the SNR and either of the selective attention effects in the pooled [$r(28) < 0.23$, $p > 0.228$] or subgroup data [$-0.33 < r(13) < 0.034$, $p > 0.229$].

IV. DISCUSSION

A. Summary of findings

The primary finding of this study is that cochlear gain is significantly reduced when participants attend to visual stimuli compared with when they attend to auditory stimuli, as indicated by the difference in SFOAE magnitudes between the AL and V attention conditions. It appears that the pattern in the overall data is driven by a subset of individuals (40%) who show large (~ 5 dB) differences in OAE magnitude between the attention conditions. The results suggest that, for these participants, changes in cochlear gain may play an important role in modulating sensory input based on attentional factors.

Interestingly, no significant difference in SFOAE magnitude was observed between conditions where attention was directed at an audio frequency remote from the SFOAE (i.e., there was no difference in emission magnitude between the AL and AH conditions). The lack of difference in emission magnitude between the two auditory conditions contrasts with the frequency specific attentional effect observed by Maison *et al.* (2001) and the clear evidence for cortical amplification of attended vs unattended auditory targets (e.g., Gutschalk *et al.*, 2008; Mesgarani and Chang, 2012). Frequency-specific effects might have been expected, based on the demonstration of frequency tuning of MOC efferent effects (Lilaonitkul and Guinan, 2012), but there also exists evidence that attention can produce broadband effects (Walsh *et al.*, 2014a, 2014b, 2015). These broadband effects

FIG. 8. (A) shows relationships between measured stimulus-frequency ear canal pressure and the magnitude of the attention effect. (B) relates the SNR to the attention effect. Dark blue symbols show the AL-V effect, while light blue symbols show the AH-V effect. Closed symbols denote participants in the original group, whereas open symbols denote participants in the replication group.

were noted in the noise within the ear canal opposite the ear where OAE evoking stimuli were presented (Walsh *et al.*, 2014a, 2014b) or by shifting the OAE analysis filter off the SFOAE frequency to examine changes in energy not evoked by the stimulus tone. Based on a recent report by Francis *et al.* (2018), it is unclear whether these broadband effects have cochlear origin and if they reflect MOC activity.

It remains to be determined whether the lack of specificity within the auditory domain is an inherent property of the attentional MOC system or whether it reflects a property of the specific task we used in this study.

B. Failure to replicate strong findings between groups

1. Group differences

Perhaps the most striking finding from this study is that the results differ strongly between two groups of participants. In the original group, more than half (9/15) of the participants showed individually significant changes in SFOAE magnitudes across attention conditions compared to only 3/15 in the replication. The groups did not differ in their performance on the task but were not matched during recruiting, so differences between groups could help explain the differences in effect.

Demographic data indicated a difference in the racial makeup of each group, with the original experimental group containing a significantly higher number of participants who identified as Asian. This difference could be important, considering some work demonstrating racial differences in OAE magnitude (McFadden *et al.*, 2012; McFadden *et al.*, 2018a; Shahnaz, 2008; Whitehead *et al.*, 1993). If racial groups differ in SFOAE magnitude, then it is possible that the SNR may also differ across racial groups, leading to differential ability to detect MOC effects on SFOAEs. We found in our sample that Asians tended to have a larger difference in SFOAE magnitude between the AL and V conditions than non-Asians, and Asians had slightly higher SNR than non-Asians. These differences were not statistically significant and therefore unlikely to be responsible for driving the differences observed between the original and replication groups.

2. Effect of ear canal pressure

There was a significant difference in overall SFOAE magnitude between the original and replication groups. We traced this difference to differing SPLs at the stimulus frequency in the ear canal while the stimulus was being delivered. Although the task, stimuli, and procedures were identical between the two groups, the equipment used for measuring OAEs was different. The original group of participants had stimuli delivered and recorded using the ER10C (Etymotic Research, Elk Grove, IL) acoustic assembly, while the replication group was tested using the ER10X system (Etymotic Research, Elk Grove, IL). While these systems were both calibrated using the same equipment, the disposable ear tips used with each system were different and may have contributed to differing SPLs in the ear canal.

To determine what effect the difference in ear canal sound pressure may have had on the overall findings, we examined the relationship between the ear canal sound pressure at the stimulus frequency and the effect of attention condition. Our correlation analysis showed no significant relationship between stimulus pressure in the ear canal and the condition effect. The lack of correlation was not due to a lack of variability in measured ear canal sound pressures in each group (see Fig. 8). We also investigated relationships between the ear canal sound pressure and other factors that are important to robust measures of MOC effects on SFOAEs. For a constant noise floor, an increase in stimulus level will result in an increase in SFOAE magnitude, and thus an increase in SNR for SFOAE measurements. Significant differences in SNR could lead to a differential ability to detect MOC-induced changes in SFOAE (Goodman *et al.*, 2013) or attentional modulations thereof. Even though SNR was significantly different between the two groups, the lack of effect seen in the replication group is unlikely to be due to a change in SNR, because there was no significant correlation between SNR and the selective attention effect (see Fig. 8). The lack of correlation was likely due to the fact that all of our participants had high SNR (>20 dB) for SFOAE measurements, as necessitated by our screening procedure, and thus attentional modulation of efferent effects, if present, should be reliably detected in all participants.

Despite the observed differences between groups in demographic and experimental data, we do not have compelling evidence that any of the factors discussed above can explain the difference between the groups in overall outcome. Below, we will focus on the combined result for all 30 participants and treat them as 1 (larger) sample.

C. Comparison to previous studies

1. Methodological improvements

Although some evidence for peripheral auditory effects of selective attention has been reported in an animal study, based on recordings of the compound action potential and cochlear microphonic (Delano *et al.*, 2007), the many previous human studies of attentional effects on peripheral auditory responses have generally failed to find robust and consistent effects of attention on cochlear responses via OAEs (Giard *et al.*, 1994; Maison *et al.*, 2001; Michie *et al.*, 1996; Srinivasan *et al.*, 2012; Srinivasan *et al.*, 2014; Walsh *et al.*, 2015). We designed the current study with several methodological manipulations with the purpose of improving our ability to detect an effect of attention on the OAE signal. The main methodological improvements included the use of SFOAEs at low stimulus levels to avoid activation of the MOCR by the test stimulus. This is important because activating the MOCR by the test stimuli may introduce interactions between two acoustically elicited efferent effects and possible attentional effects that would make the interpretation of results difficult. This is because changes in SFOAE phase due to MOCR elicited by different acoustic stimuli can be different, causing complex phase interactions that could affect the resulting SFOAE (Lilaonitkul and Guinan,

2012). An additional improvement is the use of attentional tasks in all conditions, rather than the more common “passive” comparison, for which attention cannot be properly controlled. The use of a low-level notched-noise elicitor also helps to avoid inadvertent activation of the middle ear muscle reflex (MEMR). MEMR activation produces changes in middle ear impedance that would contaminate the SFOAE measurement due to an increased or decreased energy in ear canal sound pressure at the probe frequency that could be wrongly attributed to the MOC efferent effect of interest in this study. MOCR measurements are only valid under the assumption that the middle ear impedance remains the same in the presence and absence of the MOCR elicitor. Moreover, efferent projections from higher-level auditory nuclei to the MEMR neural circuit make it possible for attention to affect the MEMR strength, and thus modulate the ear canal sound level at the OAE-evoking frequency (Mukerji *et al.*, 2010).

Previous studies used high-level, fast-rate clicks to evoke OAEs (Ferber-Viart *et al.*, 1995; Harkrider and Bowers, 2009); these have been shown to risk activation of the MOCR by probe stimuli themselves (Guinan *et al.*, 2003), potentially reducing attention-related effects. Attentional changes in distortion-product otoacoustic emissions (DPOAEs; Smith *et al.*, 2012; Srinivasan *et al.*, 2012; Srinivasan *et al.*, 2014; Wittekindt *et al.*, 2014) are difficult to interpret due to phase interactions between DPOAE components that can result in both increases and decreases in DPOAE magnitude. We also used the suppression method for estimating SFOAE, rather than estimation of the nonlinear SFOAE component (e.g., nSFOAE, Walsh *et al.*, 2015) to preserve changes that could occur linearly. In addition, our low-level (55 dB SPL) spectrally notched noise reduced the possibility of inadvertent activation of the MEMR, and our low probe level of 35 dB SPL ensured that the probe itself did not elicit measurable MOCR activity.

Most previous studies using OAEs to probe cochlear function under auditory attention have presented OAE eliciting stimuli as task-irrelevant stimuli (Avan and Bonfils, 1992; de Boer and Thornton, 2007; Ferber-Viart *et al.*, 1995; Maison *et al.*, 2001; Walsh *et al.*, 2014a, 2014b, 2015). This could lead to instances where participants tried to ignore the OAE stimuli since they could be distracting in the context of the actual task. The current study utilized SFOAE stimuli in one of the behavioral tasks to compare the cochlear function at the attended frequency to that when attention is directed to the task focused on a different frequency region. This was done to maximize the size of the attentional effect, but the results did not reveal a frequency-specific effect of attention. This finding is consistent with other studies that required selective attention to DPOAE primaries (e.g., Srinivasan *et al.*, 2012; Srinivasan *et al.*, 2014; Wittekindt *et al.*, 2014). These studies demonstrated cross-modal effects of attention, even though the frequency of DPOAE was different from the attentional target in the auditory task. Although there is anatomical evidence from animals (Brown, 1989; Liberman and Brown, 1986) and OAE evidence from humans (Lilaonitkul and Guinan, 2012; Maison *et al.*, 1997) supporting tuning of MOC effects, the current study was unable to demonstrate

that shifts in attention produce frequency-specific effects. There was a trend for lower emission magnitudes in the AH condition relative to the AL condition in the original group [see the mean bars in Fig. 5(A)], but like the AL-V difference, this effect was not reproduced in our second group of participants.

The frequency specificity of MOC effects in previous reports were attributed to changes in the stimulus properties of the MOC elicitor (Lilaonitkul and Guinan, 2012; Maison *et al.*, 1997), which were held constant in the current study. This suggests that the attentional modulation of MOC activity behaves differently than the acoustic activation of the MOC system, as attention was unable to produce an observable frequency-specific effect in this study.

Several previous reports have compared a state of attention, where participants were actively involved in some perceptual task, to a passive state of inattention (Avan and Bonfils, 1992; Puel *et al.*, 1988; Walsh *et al.*, 2014a, 2014b, 2015). These comparisons present two primary inherent difficulties. First, it is not possible to interpret the results found in these studies as strictly due to selective attention, as general arousal likely differs between an active (attention) and a passive (inattention) condition. Second, while task performance can provide some measure of how attentive participants are while completing the behavioral task, inattention does not have a corresponding behavioral metric. This means that participants could continue to attend to aspects of their environment (including OAE-evoking stimuli) during the passive task, and so any comparisons made to this condition would suffer due to the variability in attentional state across participants.

Although the methodological changes made in the current study eliminated a few potential sources of confounding factors due to the experimental design, the new design was not able to overcome the significant inter-subject variability in selective attention effects on OAE magnitude. Furthermore, this experimental design did not allow us to elucidate what is different about the participants who showed significant differences in SFOAEs between V and AL conditions. Some limitations of our design are discussed in Sec. IVC 2.

2. Limitations of the current design

Our current design is somewhat limited in its ability to relate behavioral performance to changes in SFOAE due to our choice of response method. While responding at the end of a run should reduce motion-related artifacts during the recording, it also makes it impossible to do more than a very coarse comparison between OAE magnitudes and performance. A future improvement could be to incorporate responses to targets in real time.

A second limitation comes from the fact that we used short gated stimuli to measure SFOAEs. Unlike the continuous probe tones used in most previous studies (Backus and Guinan, 2006; Lilaonitkul and Guinan, 2012, 2009), short tones make it difficult to observe slow changes due to factors such as drift in the position of the probe assembly in the ear canal. We looked for drift across conditions by comparing

both baseline SFOAE measurements and noise floor across the three attention conditions, and found no significant differences in either measure. We also looked for trends in stimulus pressure in the ear canal across each of the runs in each participant to see if there were any systematic changes. While there was some variability across the runs, we noted no systematic change in stimulus pressure in the ear canal that could produce the observed selective attention effects. This is also supported by a lack of order as a significant factor influencing emission magnitudes.

The use of a low-frequency tone for eliciting SFOAE may have limited our ability to observe MOC effects by limiting the range of cochlear gain (and potential reduction) applied to the stimulus. In animal studies, innervation of the OHCs by MOC efferents has been shown to be greater at mid-frequencies than low or high frequencies (Brown, 2011; Liberman and Brown, 1986). Physiological measurements in animals (for review, see Recio-Spinoso and Oghalai, 2017; Robles and Ruggero, 2001), as well as indirect measurements of cochlear gain in humans, also suggest that there is less gain at lower frequencies (Gorga *et al.*, 2007; Plack *et al.*, 2008). Future work utilizing the methodological improvements noted in Sec. IV B could be further improved by measuring at less apical frequencies that may have a larger dynamic range for measuring MOC effects.

3. The importance of adequate sampling

As concerns over the replicability of psychological science are currently a topic of active discussion (e.g., Open Science Collaboration, 2015), a key contribution of the current work is to highlight the need for higher power in experimental designs seeking to use SFOAEs to investigate MOC efferent effects.

Several of the more recent studies of selective attention on cochlear function (e.g., Srinivasan *et al.*, 2012; Srinivasan *et al.*, 2014; Walsh *et al.*, 2015) used relatively small ($n < 15$) samples of participants, which limited their ability to detect significant differences in emission magnitudes that were due to attention. The current study used a simple power analysis to estimate the required number of participants to observe an attention effect based on pilot data, but a sample of this size ($n = 15$) failed to secure a reproducible finding. The surprising difference in the outcome of the attention condition study between two participant samples, despite the same design and stimuli, suggests that the reported attentional effects on OAEs in humans may have been due to spurious effects in a specific sample. It is also possible that the differences between samples simply reflect large inter-subject variability in the magnitude of attentional effects, even in a group of young normally hearing listeners. Individual variability in OAE magnitudes have been explored by studies looking at biological factors like race and sex, as well as psychophysical performance (McFadden *et al.*, 2012; McFadden *et al.*, 2018a, 2018b), but future work is necessary to explore the individual variability revealed in our study.

D. Conclusions

This study provides some evidence that cochlear responses to sound may change under conditions of selective attention. Specifically, visual attention (relative to auditory attention) is accompanied by a modest reduction (~ 2.5 dB) of SFOAE magnitude, suggesting a reduction in cochlear gain of the unattended auditory stimulus. Although statistically significant for the combined pool of participants, the effect of attention condition was driven by a large significant effect in the first group of participants tested. None of the considered differences between the two participant groups provides a compelling explanation for the different outcomes. Recent studies of the attentional effects utilize sample sizes that are likely too small to reliably detect attention effects, which may explain the lack of consistent results seen in the current literature. It is possible that the efferent system is capable of modulating cochlear responses by attention, but the OAE-based measures are inadequate to detect the attentional effects because their inherent variability swamps any systematic differences related to attention. It is also possible that even with the anatomical efferent structures in place, attentional control does not extend all the way to the cochlea and instead it arises at central sites in the auditory pathways. This possibility has support from a study showing no evidence of attentional modulation in auditory-brainstem steady-state responses (Varghese *et al.*, 2015).

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grant Nos. R01 DC010374 (M.W.) and R01 DC012262 (A.J.O.). The authors declare no competing financial interests.

- Avan, P., and Bonfils, P. (1992). "Analysis of possible interactions of an attentional task with cochlear micromechanics," *Hear. Res.* **57**, 269–275.
- Backus, B. C., and Guinan, J. J. (2006). "Time-course of the human medial olivocochlear reflex," *J. Acoust. Soc. Am.* **119**, 2889–2904.
- Brown, M. C. (1989). "Morphology and response properties of single olivocochlear fibers in the guinea pig," *Hear. Res.* **40**, 93–109.
- Brown, M. C. (2011). "Anatomy of olivocochlear neurons," in *Audit. Vestib. Efferents*, edited by D. K. Ryugo and R. R. Fay (Springer, New York), pp. 17–37.
- Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.* **25**, 975–979.
- Choi, I., Rajaram, S., Varghese, L. A., and Shinn-Cunningham, B. G. (2013). "Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography," *Front. Hum. Neurosci.* **7**, 115.
- de Boer, J., and Thornton, A. R. D. (2007). "Effect of subject task on contralateral suppression of click evoked otoacoustic emissions," *Hear. Res.* **233**, 117–123.
- Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (2007). "Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas," *J. Neurosci.* **27**, 4146–4153.
- Ding, N., and Simon, J. Z. (2012). "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *J. Neurophysiol.* **107**, 78–89.
- Ferber-Viart, C., Duclaux, R., Collet, L., and Guyonnard, F. (1995). "Influence of auditory stimulation and visual attention on otoacoustic emissions," *Physiol. Behav.* **57**, 1075–1079.
- Francis, N. A., Zhao, W., and Guinan, J. J., Jr. (2018). "Auditory attention reduced ear-canal noise in humans by reducing subject motion, not by medial olivocochlear efferent inhibition: Implications for measuring otoacoustic emissions during a behavioral task," *Front. Syst. Neurosci.* **12**, 42 (published online).

- Frøehlich, P., Collet, L., Chanal, J.-M., and Morgon, A. (1990). "Variability of the influence of a visual task on the active micromechanical properties of the cochlea," *Brain Res.* **508**, 286–288.
- Frøehlich, P., Collet, L., and Morgon, A. (1993). "Transiently evoked otoacoustic emission amplitudes change with changes of directed attention," *Physiol. Behav.* **53**, 679–682.
- Giard, M.-H., Collet, L., Bouchet, P., and Pernier, J. (1994). "Auditory selective attention in the human cochlea," *Brain Res.* **633**, 353–356.
- Gilbert, A. N., and Wysocki, C. J. (1992). "Hand preference and age in the United States," *Neuropsychologia* **30**, 601–608.
- Goodman, S. S., Mertes, I. B., Lewis, J. D., and Weissbeck, D. K. (2013). "Medial olivocochlear-induced transient-evoked otoacoustic emission amplitude shifts in individual subjects," *J. Assoc. Res. Otolaryngol.* **14**, 829–842.
- Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann B. (2007). "Low-frequency and high-frequency cochlear nonlinearity in humans," *J. Acoust. Soc. Am.* **122**, 1671–1680.
- Guinan, J. J., Jr. (2006). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," *Ear Hear.* **27**, 589–607.
- Guinan, J. J., Jr. (2010). "Cochlear efferent innervation and function," *Curr. Opin. Otolaryngol. Head Neck Surg.* **18**, 447–453.
- Guinan, J. J., Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," *J. Assoc. Res. Otolaryngol.* **4**, 521–540.
- Gutschalk, A., Micheyl, C., and Oxenham, A. J. (2008). "Neural correlates of auditory perceptual awareness under informational masking," *PLoS Biol.* **6**, 1156–1165.
- Harkrider, A. W., and Bowers, C. D. (2009). "Evidence for a cortically mediated release from inhibition in the human cochlea," *J. Am. Acad. Audiol.* **20**, 208–215.
- Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.
- Khalfa, S., Vuillet, E., and Collet, L. (1998). "Influence of handedness on peripheral auditory asymmetry," *Eur. J. Neurosci.* **10**, 2731–2737.
- Kruschke, J. K. (2013). "Bayesian estimation supersedes the t test," *J. Exp. Psychol. Gen.* **142**, 573–603.
- Liberman, M. C., and Brown, M. C. (1986). "Physiology and anatomy of single olivocochlear neurons in the cat," *Hear. Res.* **24**, 17–36.
- Lilaonitkul, W., and Guinan, J. J. (2009). "Human medial olivocochlear reflex: Effects as functions of contralateral, ipsilateral, and bilateral elicitor bandwidths," *J. Assoc. Res. Otolaryngol.* **10**, 459–470.
- Lilaonitkul, W., and Guinan, J. J. (2012). "Frequency tuning of medial-olivocochlear-efferent acoustic reflexes in humans as functions of probe frequency," *J. Neurophysiol.* **107**, 1598–1611.
- Maison, S., Micheyl, C., and Collet, L. (1997). "Medial olivocochlear efferent system in humans studied with amplitude-modulated tones," *J. Neurophysiol.* **77**, 1759–1768.
- Maison, S., Micheyl, C., and Collet, L. (2001). "Influence of focused auditory attention on cochlear activity in humans," *Psychophysiology* **38**, 35–40.
- McFadden, D., Pasanen, E. G., Leshikar, E. M., Hsieh, M. D., and Maloney, M. M. (2012). "Comparing behavioral and physiological measures of combination tones: Sex and race differences," *J. Acoust. Soc. Am.* **132**, 968–983.
- McFadden, D., Pasanen, E. G., Maloney, M. M., Leshikar, E. M., and Pho, M. H. (2018a). "Correlations between otoacoustic emissions and performance in common psychoacoustical tasks," *J. Acoust. Soc. Am.* **143**, 2355–2367.
- McFadden, D., Pasanen, E. G., Maloney, M. M., Leshikar, E. M., and Pho, M. H. (2018b). "Differences in common psychoacoustical tasks by sex, menstrual cycle, and race," *J. Acoust. Soc. Am.* **143**, 2338–2354.
- Mesgarani, N., and Chang, E. F. (2012). "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature* **485**, 233–236.
- Michie, P. T., LePage, E. L., Solowij, N., Haller, M., and Terry, L. (1996). "Evoked otoacoustic emissions and auditory selective attention," *Hear. Res.* **98**, 54–67.
- Mukerji, S., Windsor, A. M., and Lee, D. J. (2010). "Auditory brainstem circuits that mediate the middle ear muscle reflex," *Trends Amplif.* **14**, 170–191.
- O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor E. C. (2015). "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cereb. Cortex* **25**, 1697–1706.
- Open Science Collaboration. (2015). "Estimating the reproducibility of psychological science," *Science* **349**, aac4716.
- Penner, M. J., Glotzbach, L., and Huang, T. (1993). "Spontaneous otoacoustic emissions: Measurement and data," *Hear. Res.* **68**, 229–237.
- Picton, T. W., and Hillyard, S. A. (1974). "Human auditory evoked potentials II: Effects of attention," *Electroencephalogr. Clin. Neurophysiol.* **36**, 191–199.
- Picton, T. W., Hillyard, S. A., Galambos, R., and Schiff, M. (1971). "Human auditory attention: A central or peripheral process?," *Science* **173**, 351–353.
- Plack, C. J., Oxenham, A. J., Simonson, A. M., O'Hanlon, C. G., Drga, V., and Arifanto, D. (2008). "Estimates of compression at low and high frequencies using masking additivity in normal and impaired ears," *J. Acoust. Soc. Am.* **123**, 4321–4330.
- Puel, J.-L., Bonfils, P., and Pujol, R. (1988). "Selective attention modifies the active micromechanical properties of the cochlea," *Brain Res.* **447**, 380–383.
- Puria, S., Guinan, J. J., and Liberman, M. C. (1996). "Olivocochlear reflex assays: Effects of contralateral sound on compound action potentials versus ear-canal distortion products," *J. Acoust. Soc. Am.* **99**, 500–507.
- Recio-Spinoso, A., and Oghalai, J. S. (2017). "Mechanical tuning and amplification within the apex of the guinea pig cochlea," *J. Physiol.* **595**, 4549–4561.
- Robles, L., and Ruggero, M. A. (2001). "Mechanics of the mammalian cochlea," *Physiol. Rev.* **81**, 1305–1352.
- Shahnaz, N. (2008). "Transient evoked otoacoustic emissions (TEOAEs) in Caucasian and Chinese young adults," *Int. J. Audiol.* **47**, 76–83.
- Shera, C. A., and Guinan, J. J. (1999). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," *J. Acoust. Soc. Am.* **105**, 782–798.
- Smith, D. W., Aouad, R. K., and Keil, A. (2012). "Cognitive task demands modulate the sensitivity of the human cochlea," *Front. Psychol.* **3**, 30.
- Srinivasan, S., Keil, A., Stratis, K., Osborne, A. F., Cerwonka, C., Wong, J., Rieger, B. L., Polcz, V., and Smith, D. W. (2014). "Interaural attention modulates outer hair cell function," *Eur. J. Neurosci.* **40**, 3785–3792.
- Srinivasan, S., Keil, A., Stratis, K., Woodruff Carr, K. L., and Smith, D. W. (2012). "Effects of cross-modal selective attention on the sensory periphery: Cochlear sensitivity is altered by selective attention," *Neuroscience* **223**, 325–332.
- Varghese, L., Bharadwaj, H. M., and Shinn-Cunningham, B. G. (2015). "Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses," *Brain Res.* **1626**, 146–164.
- Vuillet, E., Collet, L., and Duclaux, R. (1991). "Effect of contralateral acoustic stimulation on active cochlear micromechanical properties in human subjects: Dependence on stimulus variables," *J. Neurophysiol.* **65**, 724–735.
- Walsh, K. P., Pasanen, E. G., and McFadden, D. (2014a). "Selective attention reduces physiological noise in the external ear canals of humans I: Auditory attention," *Hear. Res.* **312**, 143–159.
- Walsh, K. P., Pasanen, E. G., and McFadden, D. (2014b). "Selective attention reduces physiological noise in the external ear canals of humans II: Visual attention," *Hear. Res.* **312**, 160–167.
- Walsh, K. P., Pasanen, E. G., and McFadden, D. (2015). "Changes in otoacoustic emissions during selective auditory and visual attention," *J. Acoust. Soc. Am.* **137**, 2737–2757.
- Whitehead, M. L., Kamal, N., Lonsbury-Martin, B. L., and Martin, G. K. (1993). "Spontaneous otoacoustic emissions in different racial groups," *Scand. Audiol.* **22**, 3–10.
- Wittekindt, A., Kaiser, J., and Abel, C. (2014). "Attentional modulation of the inner ear: A combined otoacoustic emission and EEG study," *J. Neurosci.* **34**, 9995–10002.