Changes in otoacoustic emissions during selective auditory and visual attention

Kyle P. Walsh,[a] Edward G. Pasanen, and Dennis McFadden

Department of Psychology and Center for Perceptual Systems, University of Texas,
1 University Station A8000, Austin, Texas 78712-0187, USA

Previous studies have demonstrated that the otoacoustic emissions (OAEs) measured during behavioral tasks can have different magnitudes when subjects are attending selectively or not attending. The implication is that the cognitive and perceptual demands of a task can affect the first neural stage of auditory processing—the sensory receptors themselves. However, the directions of the reported attentional effects have been inconsistent, the magnitudes of the observed differences typically have been small, and comparisons across studies have been made difficult by significant procedural differences. In this study, a nonlinear version of the stimulus-frequency OAE (SFOAE), called the nSFOAE, was used to measure cochlear responses from human subjects while they simultaneously performed behavioral tasks requiring selective auditory attention (dichotic or diotic listening), selective visual attention, or relative inattention. Within subjects, the differences in nSFOAE magnitude between inattention and attention conditions were about 2–3 dB for both auditory and visual modalities, and the effect sizes for the differences typically were large for both nSFOAE magnitude and phase. These results reveal that the cochlear efferent reflex is differentially active during selective attention and inattention, for both auditory and visual tasks, although they do not reveal *how* attention is improved when efferent activity is greater.
© 2015 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4919350]

[CAS] Pages: 2737–2757

I. INTRODUCTION

Beginning soon after the first reports on the existence of an olivo-cochlear bundle (OCB) of neurons connecting the hindbrain to the cochlea (Rasmussen, 1946, 1953), speculation began about whether this efferent circuit somehow might be involved in modulating the flow of auditory afferent information under conditions of selective attention (e.g., Hernández-Peón *et al.*, 1956). Such speculation only increased as knowledge grew about the extensive network of efferent circuits that exists above the hindbrain (Mulders and Robertson, 2000a,b). Accordingly, numerous studies have been done over the years aimed at determining whether, and how, the efferent system behaves under conditions of selective attention. Many of these studies have used otoacoustic emissions (OAEs) as their dependent measure. OAEs have a number of advantages: They reveal response characteristics of the active cochlea, they are sensitive to changes in response due to olivo-cochlear efferent feedback, and they are a noninvasive physiological measure. A common result of these studies is an apparent increase in efferent activity during selective attention (e.g., Ferber-Viart *et al.*, 1995; Froehlich *et al.*, 1993; Maison *et al.*, 2001; Puel *et al.*, 1988; Walsh *et al.*, 2014a,b). However, Harkrider and Bowers (2009) reported that suppression of click-evoked OAEs (CEOAEs) by a contralateral noise was *reduced* during

selective auditory attention compared to their inattention condition.

Here we report two studies on the effects of selective attention on cochlear responses. Our primary measure was a form of stimulus-frequency otoacoustic emission (SFOAE) that we call the nonlinear SFOAE, or nSFOAE (Walsh *et al.*, 2010a,b). The nSFOAE is a measure of how much SFOAE magnitude deviates from linear growth as stimulus level is increased. This measure was obtained during blocks of trials when subjects were, or were not, required to attend to sequences of digits presented either aurally or visually. Additional auditory stimuli known to activate the medial olivo-cochlear (MOC) component of the OCB were interleaved with the digit stimuli within each trial in order to elicit the nSFOAE response. Thus, the nSFOAE measures reported here were obtained contemporaneously with the presentation of the attended digits. For most subjects, the magnitudes of the nSFOAEs were different under attention and inattention, revealing that the MOC efferent system was differentially active under the two conditions. These results are similar to those reported previously from the same studies, using the same subjects (Walsh *et al.*, 2014a,b): Those previous reports concentrated on the nSFOAE responses obtained during a brief silent period *following* each eliciting stimulus. Here we concentrate on the nSFOAE responses obtained *during* the presentation of the SFOAE-eliciting stimuli, what we call the perstimulatory responses.

Previous studies using OAEs to study attention typically used relatively simple behavioral tasks. For example, Puel *et al.* (1988) and Froehlich *et al.* (1990) had their subjects count the number of times a target letter appeared on a

---

screen, Froehlich *et al.* (1993), Ferber-Viart *et al.* (1995), Meric and Collet (1994) measured reaction times following a light flash or a tone pip, and Giard *et al.* (1994), Maison *et al.* (2001), and Harkrider and Bowers (2009) asked their subjects simply to attend to sounds—sometimes those used to evoke their OAE measure—presented to one ear. By contrast, the present studies used more demanding behavioral tasks, and also used different acoustic stimuli to evoke the OAE. All previous investigations have used transient stimuli—either brief tones or clicks—to evoke the OAE, but these stimuli appear not to have been the best choice for studying the effects of the MOC efferent system. Guinan *et al.* (2003) reported that the stimuli used commonly as "probes" of efferent activation—clicks, tone pips, and tone pairs—often can themselves contribute inadvertently as "elicitors" of efferent activity. This means that in those previous studies, the efferent system could have been activated to some degree whether or not subjects were attending, and whether or not a deliberate MOC-eliciting stimulus was presented simultaneously with the OAE-evoking (and perhaps also MOC-eliciting) stimulus. This is a problematic possibility.

In contrast, Guinan *et al.* (2003) reported that long-duration tones of moderate intensity, like those used to evoke stimulus-frequency OAEs (SFOAEs), are relatively ineffective elicitors of the efferent system. Accordingly, our nSFOAE measure was elicited by a long-duration tone presented both alone and with a wideband noise (which *is* known to activate the MOC reflex). The behavioral tasks assigned the subjects did not pertain to these stimuli but to other stimuli that were interleaved temporally with them. In the first study, subjects attended to series of spoken digits, and in the second study, they attended to series of visually presented digits. In both studies, the individual digits alternated with the nSFOAE stimuli, meaning that our measures of the state of the cochlea were contemporaneous with the processes of attention demanded of the subject. The results obtained showed that the sounds we used to evoke emissions from the cochlea were processed differentially depending on the attentional demands imposed upon the listener, and this was true whether auditory or visual attention was required.

II. METHODS

A. General

Two studies are described here, one involving attention to sequences of auditory stimuli and the other to sequences of visual stimuli. The auditory study was completed before the beginning of the visual study, and the same subjects participated in both studies (except for the loss of one subject at the end of the auditory study). The procedures for collecting and analyzing the nSFOAE responses were essentially identical for the two studies. The differences in the behavioral tasks are described below.

The Institutional Review Board at The University of Texas at Austin approved the procedures described here, all subjects provided their informed consent prior to any testing, and the subjects were paid for their participation.

Supplementary materials for this article can be found online.[1]

B. Subjects

Two males (both aged 22) and six females (aged 20–25) participated in the auditory study; one female was not available to complete the visual study. All subjects had normal hearing [$\leq$15 dB Hearing Level (HL)] at octave frequencies between 250 and 8000 Hz, and normal middle-ear and tympanic reflexes, as determined using an audiometric screening device (Auto Tymp 38, GSI/VIASYS, Inc., Madison, WI). The test of the middle-ear reflex (MER) involved individually presenting tones at four frequencies to the test ear, each at as many as three levels, while the compliance of the middle-ear system was measured. For the 0.5- and 4.0-kHz tones, the test levels were 80, 90, and 100 dB HL, and for the 1.0- and 2.0-kHz tones, the test levels were 85, 95, and 105 dB HL. All tone series began with the weakest stimulus and increased in level only if an MER was not observed. An MER was defined as a decrease in the compliance of the middle-ear ossicles of $0.05\,cm^3$ equivalent volume or more during the presentation of the test stimulus. At 4.0 kHz, a frequency used to elicit the nSFOAE response in our studies, only three subjects had an MER in either the left or the right ear at 80 dB HL, the lowest level tested. The level of the 4.0-kHz tone used in our attention studies was only 60 dB sound-pressure level (SPL), but we cannot be certain that it was incapable of activating the MER in all subjects. No subject had a spontaneous otoacoustic emission (SOAE) stronger than -15.0 dB SPL within 600 Hz of the 4.0-kHz tone used to elicit the nSFOAE response.

Each subject was tested individually while seated in a reclining chair inside a double-walled, sound-attenuated room. Two insert earphone systems (described below) delivered sound directly to the external ear canals. A computer monitor attached to an articulating mounting arm was positioned at a comfortable viewing distance from the subject; it was used to provide instructions, stimuli, response alternatives, and trial-by-trial feedback for the auditory and visual behavioral tasks. A numeric keypad was provided to the subject for indicating his or her behavioral responses. The experimenter (author K.P.W.) sat adjacent to the subject, in front of a computer screen from which he conducted the test session.

C. Behavioral measures

1. Selective auditory-attention conditions

For the auditory study, there were two selective-attention conditions: One required dichotic listening and the other diotic listening. For the dichotic-listening condition, two competing speech streams were presented separately to the ears, and the task of the subject was to attend to one of the speech streams. In one ear, the talker was female, in the other ear the talker was male, and the ear receiving the female talker was selected trial by trial from a closed set of random permutations. Within each block of trials, the female voice was in the right ear approximately equally as often as

in the left ear. On each trial, the two talkers simultaneously spoke two different sequences of seven numerical digits. Each digit was selected randomly with replacement (0–9), and the digit sequence spoken by the single female talker was selected independently from that spoken by the single male talker. Digits were presented during 500-ms intervals that were separated by 330-ms interstimulus intervals (ISIs). As described more fully below, the stimulus waveforms used to elicit the nSFOAE response were presented in those six ISIs between the seven digits. At the end of each trial in those blocks requiring attention, the subject was shown two sets of five digits (identical except for one digit) and was allowed 2000 ms to indicate with an appropriate key press which set corresponded to the middle five digits spoken by the female voice. Visual trial-by-trial feedback as to the correctness of these choices was provided on the subject's computer screen. For the visual-attention study, the trial-timing sequence was the same, but the digits were presented visually, not aurally (see Sec. IV below).

The diotic-listening condition was similar to the dichotic-listening condition with the exception that the male and female voices were presented simultaneously to both ears on each trial, not to separate ears. Thus, the dichotic-listening condition required attention to one of two spatial locations—the left or the right ear—whereas the diotic-listening condition required subjects to disambiguate two speech streams that seemed to originate from the same location in space, roughly in the center of the head. The timing of each diotic trial was identical to each dichotic trial.

In order to assess the effects of attention, control conditions not requiring attention were needed. Ideally, all aspects of the task and stimuli ought to be identical to those in the attention condition, the only difference being that the subject was not required to attend. The problem with such an arrangement here is that speech is a unique stimulus for humans, and it is not clear that people can truly ignore speech when they hear it. Accordingly, for our primary inattention conditions in both the auditory- and visual-attention studies, the speech waveforms were replaced with sounds having spectral and temporal characteristics like speech (called speech-shaped noises or SSNs; details are below), but which were not intelligible. The individual SSNs sounded different from each other but did not sound like the digits from which they were derived. Thus, all the attention and inattention conditions contained stimuli having similar spectral and temporal characteristics (one exception existed in the visual-attention study), so that whatever physiological mechanisms were reflexively activated in the attention conditions also were activated in the inattention conditions.

Because our primary dependent variable was a physiological measure that can be sensitive to motor activity, the inattention control condition needed a motor response like that required in the attention conditions. Accordingly, for those blocks of trials *not* requiring attention to the speech sounds, the subject pressed a key after the last of the seven stimuli on each trial. Thus, some attention *was* required of the subject (to the trial-timing sequence), just no attention to the content of the speech prior to making a demanding discrimination at the end of each trial.

For all trials in all blocks, the time required for the subject to make the behavioral response was recorded; for the inattention blocks, this reaction time (RT) was the sole behavioral measure obtained. For all conditions, the RT clock began at the end of the seventh 500-ms epoch containing speech (or SSN) and was terminated by a key press. If no behavioral response occurred prior to the end of the 2000-ms response interval, neither the behavioral data nor the physiological data from that trial were saved. Trials with no response occurred infrequently. In the attention conditions, reaction times and percent correct values were recorded and averaged separately for correct and incorrect trials.

Nominally, all test trials were run in blocks of 30. However, whenever a subject failed to respond on the behavioral task, or the physiological response failed to meet pre-established criteria, that trial was discarded, and another trial was added to the end of the block (the visual display indicated this to the subject). Accordingly, the duration of a block of trials ranged between 4 and 6 min, and the behavioral measures (percent correct judgments and RT) necessarily were based on slightly different numbers of trials across conditions and subjects.

On each block of trials requiring attention to the digit strings, the percentage of correct responses was calculated automatically, and saved to disk. Those values were averaged over four to six repetitions of the same condition to yield an overall estimate of performance for that test condition. When designing these studies, the goal was to have behavioral attention tasks that were cognitively challenging for most subjects, but not too difficult for any of them. Specifically, we wanted everyone to perform above about 70% correct, but below 100% correct, and as will be seen, we were successful at this. Because there was no measure of correctness on the inattention blocks for comparison with the attention blocks, there never was any intention for these behavioral measures to be crucial to the primary question of interest (whether activation of the MOC efferent system differed between selective attention and inattention). All we needed was evidence that the subjects were attending to the digit sequences, and that evidence consisted of good performance on the discrimination tasks. Furthermore, the fact that performance was typically less than perfect led us to conclude that the attention task did require cognitive effort.

2. Speech stimuli

One female talker and one male talker were used to create the speech stimuli. Their speech was recorded using the internal microphone on an iMac computer that was running the Audacity (Sourceforge, Slashdot Media, San Francisco, CA) (audacity.sourceforge.net) software application. Each recording was the entire sequence of ten digits (0–9), spoken slowly, and several recordings were obtained from each of the two talkers. After recording, the most neutral-sounding sequence from each talker was selected. The individual digit waveforms then were cropped from the selected recordings and fitted to a 500-ms window by aligning the onset of each waveform with the onset of the window, and by adding the appropriate number of zero-amplitude samples ("zeros") to

J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh *et al.*: Efferents and attention   2739

the end of each waveform to fill the window. The recordings were made using a 50-kHz sampling rate and 16-bit resolution, and they were not filtered or processed further before being saved individually to disk.

At the start of an experimental condition, the 20 selected digits (ten from the female talker and ten from the male talker) were read into a custom-written LabVIEW (National Instruments, Austin, TX) program once, and then held in memory for the duration of the condition. Before presentation, all waveforms were lowpass filtered at 3.0 kHz, and they were equalized in level using the following procedure. The mean root-mean-square (rms) amplitude of the ten digits spoken by the female talker was calculated, and then each of the 20 waveforms was scaled by a fixed amount such that the overall level of each waveform was about 50 dB SPL. This operation was performed on the spoken digits and SSN stimuli alike.

a. Speech-shaped noise (SSN) stimuli. Each of the 20 speech waveforms used in the auditory-attention (and the visual-attention) studies was transformed into a corresponding SSN stimulus as follows: For each digit, a Fast Fourier Transform (FFT) was taken of the entire utterance and used to create a 500-ms sample of noise having the same long-term spectral characteristics as the spoken digit. The envelope of each spoken digit was extracted using a Hilbert transform, the envelope was lowpass filtered at 500 Hz, and the resulting envelope function was applied to the relevant sample of noise. The resulting sounds were not intelligible as speech, but the pitches of the SSNs derived from the female talker were noticeably higher than those from the male talker. When presenting strings of SSN stimuli, the same rules about sampling and sequencing were used as with the real speech sounds. Subjects were never tested on the *content* of the SSNs, but in some inattention conditions, they did have to attend to the *timing* of the SSNs in order to respond that the digit sequence was completed.

D. Physiological measures

As noted, the subjects were tested individually while comfortably seated in a commercial sound room. The nSFOAE-eliciting stimuli (and the speech sounds) were delivered directly to the ears by two insert earphone systems. For the *right ear*, two Etymotic ER-2 earphones (Etymotic, Elk Grove Village, IL) were attached to plastic sound-delivery tubes that were connected to an ER-10A microphone capsule. The microphone capsule had two sound-delivery ports that were enclosed by the foam ear-tip that was fitted in the ear canal. The nSFOAE responses were elicited by sounds presented via the ER-2 earphones and recorded using the ER-10A microphone. For the *left ear*, one ER-2 earphone presented the stimuli to the external ear canal via a plastic tube passing through a foam ear-tip; there was no microphone in the left ear. That is, the nSFOAE stimuli were presented to both ears simultaneously, but nSFOAE responses were recorded from the right ear only.

The acoustic stimuli (speech sounds and nSFOAE-eliciting sounds) and the nSFOAE responses all were digitized using a National Instruments sound board (PCI-MIO-16XE-10) installed in a Macintosh G4 computer, and stimulus presentation and nSFOAE recording both were implemented using custom-written LabVIEW software (National Instruments, Austin, TX). The sampling rate for both input and output was 50 kHz with 16-bit resolution. The stimulus waveforms were passed from the digital-to-analog converter in the sound board to a custom-built, low-power amplifier before being passed to the earphones for presentation. The analog output of the microphone was passed to an Etymotic preamplifier (20 dB gain), and then to a custom-built amplifier/bandpass filter (14 dB gain, filtered from 0.4 to 15.0 kHz), before being passed to the analog-to-digital converter in the sound board.

1. The nSFOAE procedure

The physiological measure used here was a nonlinear version of the stimulus-frequency otoacoustic emission (SFOAE), called the nSFOAE (Walsh *et al.*, 2010a, 2014a). The SFOAE is a continuous tonal reflection that is evoked from the cochlea by a long-duration tone (Kemp, 1980); after a time delay, the reflection adds with the input sound in the ear canal.

Because the SFOAE is so weak, it is difficult to extract from the sound in the ear canal, dominated as it is by the stimulus tone. Our procedure for extracting the nSFOAE response from the ear-canal sound is based upon the "double-evoked" procedure described by Keefe (1998). At the heart of this procedure are three acoustic presentations (a "triplet") using the two earphones in the right ear. For the first presentation, the SFOAE-evoking stimulus is delivered to the ear via one of the ER-2 earphones at a specified SPL. For the second presentation, the exact same stimulus is delivered via the other ER-2 earphone at that same level. For the third presentation, both earphones deliver the exact same stimuli simultaneously and in phase; accordingly, the level in the ear canal is approximately 6 dB greater than the levels in each of the first two presentations. For each presentation of each triplet, the sound in the ear canal (called the response) is recorded and saved. Then the responses from the first two presentations are summed point-for-point in time and the response to the third presentation is subtracted from that summed response. To the extent that the presentation and recording systems are linear, the remaining "difference waveform" is devoid of the original stimulus. The nSFOAE can be thought of as the ongoing magnitude of the failure of additivity (Walsh *et al.*, 2010a). To obtain a stable estimate of the nSFOAE, the difference waveforms were summed across many triplets from the same block of trials.

If only linear processes were involved, then when the response to the third presentation is subtracted from the sum of the responses to the first two presentations, the resulting nSFOAE would be close to the noise floor of the measurement system. However, in our studies, the result always is a difference waveform whose magnitude is substantially above the noise floor. The reason is that the response to the third presentation of each triplet always is weaker than the sum of

the responses to the first two presentations, in accord with the known compressive characteristic of the normal human cochlea over moderate SPLs (e.g., Cooper, 2004). As a check on the software and the calibrations, the same SFOAE-evoking stimuli were presented to a passive cavity (a 0.5-cc syringe) instead of a human ear, using exactly the same procedures and equipment as used with humans, and the result of the subtraction of responses *was* an essentially perfect cancellation. That difference waveform was comparable in magnitude to other estimates of the noise floor of our OAE-measurement system, at about −13.0 dB SPL (at 4.0 kHz).

The stimulus used here to elicit the nSFOAE always was a long-duration tone presented in wideband noise. The tone was 4.0 kHz, 300 ms in duration, and had a level of 60 dB SPL. The noise had a bandwidth of 0.1–6.0 kHz, was 250 ms in duration, and had an overall level of about 62.7 dB SPL (a spectrum level of about 25 dB, so the tone-to-noise ratio was about 35 dB). The onset of the tone always preceded the onset of the noise by 50 ms. The tone was gated using a 5-ms cosine-squared rise and decay, and the noise was gated using a 2-ms cosine-squared rise and decay. The same random sample of noise was used across all presentations of a triplet, across all triplets, and across all subjects for both the auditory- and visual-attention studies. (The use of a single sample of frozen noise was important for the purpose of averaging nSFOAE responses across repeated conditions, and for comparing these averaged responses across the various experimental conditions.)

The digit stimuli and the SFOAE-eliciting stimuli were interleaved in time, as shown for a single dichotic-attention trial in Fig. 1. The two traces show the stimuli presented separately to the two ears, the female voice to one ear and the male voice to the other ear. Note that the interleaved SFOAE-eliciting stimuli were simultaneous and identical in the two ears (diotic), and the digits were simultaneous but different (dichotic). Because there were six SFOAE-eliciting presentations per trial, estimates for two triplets were obtained from each trial (the third and sixth presentations always were double amplitude, as shown). During data collection and data analysis, the estimates for the two triplets were kept separate; the responses themselves were averaged across trials within a block, and response magnitudes were averaged across blocks (see below), but no data were combined across triplets. On every trial, a silent 2000-ms response interval, and then a 200-ms feedback interval, followed the seventh digit stimulus. Thus, a single trial of the nSFOAE procedure lasted approximately 7 s. As Fig. 1 makes clear, during the auditory dichotic-attention blocks, about half of the trials produced nSFOAE responses from the right ear when attention was directed to the right ear (called the "ipsilateral" condition), and about half of the trials produced nSFOAE responses from the right ear when attention was directed to the left ear (the "contralateral" condition). These responses were kept separate initially. For the diotic-attention condition, there was no ipsilateral/contralateral distinction. Shown in Fig. 1 are small white boxes at the end of each SFOAE-eliciting stimulus. These mark the 30-ms silent period that was the topic of our previous reports (Walsh *et al.*, 2014a,b).

Each block of trials began with a calibration period, during which no speech stimuli were presented. For calibration, the level of a 500-Hz tone was adjusted in the right ear canal of the subject to attain 65 dB SPL. This routine was run separately for each of the ER-2 earphones. The calibration factors obtained were used to scale the amplitude of the
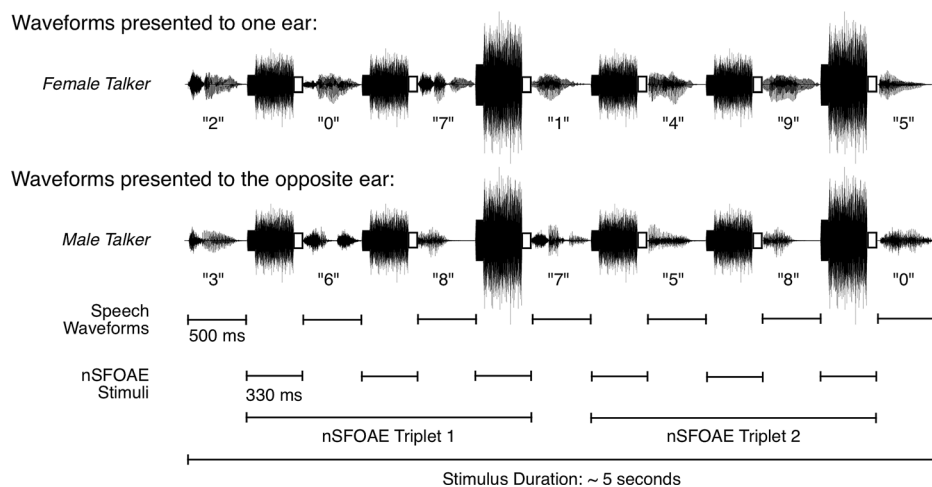

FIG. 1. Schematic showing how the speech sounds and the nSFOAE-eliciting stimuli were interleaved during one trial of the dichotic-listening condition in the auditory study. Each ear was presented with a series of seven spoken digits, one series spoken by a female talker, and the other series spoken simultaneously by a male talker. The ear receiving the female talker was selected randomly on each trial. Each digit was presented in a 500-ms temporal window. A 330-ms ISI separated consecutive digits, during which the nSFOAE-eliciting stimuli were presented. The latter always was composed of a 300-ms tone and a 250-ms frozen sample of wideband noise, and the onset of the tone always preceded the onset of the noise by the difference in their durations. A 30-ms silent period, shown here as an open rectangle, followed each nSFOAE-eliciting stimulus for the purpose of estimating the magnitude of the physiological noise in the nSFOAE recordings. The nSFOAE cancellation procedure was performed separately on each of the two triplets presented on each trial, yielding two estimates of the nSFOAE per trial. Although not shown here, a 2000-ms silent response interval and a 200-ms feedback interval completed each trial. During the response interval, the subject performed a two-alternative matching task based on the digits spoken by the female talker. For each block of trials, the physiological responses from the trials having a correct behavioral response were based on about 20 to 30 trials. [Reprinted with permission from Walsh *et al.* (2014a).]

J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh *et al.*: Efferents and attention   2741

stimulus used to elicit the nSFOAE so that the presentation levels for the two earphones were as similar as possible. This calibration routine was followed by two criterion-setting routines, and the main data-acquisition routine. The first criterion-setting routine was based on the responses obtained during the first 12 trials of each block; during those 12 trials, all nSFOAE responses were accepted (for both triplets) unless the peak amplitude of the response was greater than 45 dB SPL. (Waveforms whose amplitudes exceeded this limit typically were observed when the subject moved, swallowed, or produced some other artifactual noise.) All of the accepted nSFOAE responses collected during the first 12 trials were averaged point-by-point, and the resultant waveform served as the foundation for the accumulating nSFOAE average to be constructed during the main acquisition routine. Furthermore, the rms value of each accepted nSFOAE response was computed, and this distribution of rms values was used to evaluate subsequent responses during the main acquisition routine. The second criterion-setting routine consisted of a 20-s recording in the quiet during which no sound was presented to the ears; this was done just prior to every block. The median rms voltage from this recording was calculated, and was used as a measure of the ambient (physiological) noise level of that individual subject for that block of trials.

During actual data collection, each new nSFOAE response was compared to the responses collected during the two criterion-setting routines and was accepted into the accumulating nSFOAE average if *either* one of the two criteria was satisfied. First, the rms value of the new nSFOAE was compared to the distribution of rms values collected during the first criterion-setting routine. If the new rms value was less than 0.25 standard deviations above the median rms of the saved distribution, then the new nSFOAE response was added to the accumulating average. Second, each new nSFOAE response was subtracted point-for-point from the accumulating nSFOAE average. The rms of this difference waveform was computed, then converted to dB SPL. If the magnitude of the difference waveform was less than 6.0 dB SPL above the noise level measured earlier in the quiet, the new nSFOAE was accepted into the accumulating average. Each nSFOAE always was evaluated by both criteria, but only one criterion needed to be satisfied for the data to be accepted for averaging. The nSFOAE waveforms accepted from triplet 1 of a trial always were evaluated and averaged separately from the nSFOAEs evaluated and averaged from triplet 2 of a trial. At the end of each trial, subjects were given feedback on the computer screen as to which nSFOAE waveforms (triplet 1 or 2) were accepted. The block of trials terminated when behavioral responses had been given to at least 30 trials on which the nSFOAE response had been accepted for both triplets 1 and 2 (including trials with incorrect behavioral responses). This typically required about 35–50 trials, or about 4–6 min.

2. Subject training

As noted above, each subject was screened for normal hearing prior to participating in the studies described here.

After the screening procedure, the subjects who passed were given immediate training with the simultaneous behavioral and physiological stimuli by completing several blocks of trials of the auditory-inattention listening condition. During this training, subjects became familiar with the general testing and recording procedure. They learned to remain as still as possible during the test blocks and to press the response key without head or other body movements, permitting artifact-free OAE recordings. The subjects were not given training with the selective auditory-attention conditions prior to data collection. However, the two male subjects (L01 and L05) participated in an earlier, pilot version of this experiment that used the same experimental conditions, but slightly different control conditions, so they *were* experienced with the auditory-attention conditions.

E. Data analysis

At the end of each block of trials for the *diotic* auditory-attention condition, there were four physiological measures: An averaged physiological response for all the trials having correct behavioral responses, another average for those having incorrect behavioral responses, and each of those separately for triplets 1 and 2. For the *dichotic* auditory-attention condition, there were eight physiological measures at the end of each block of trials because the four measures described for the diotic condition were kept separately for trials on which the female voice (or female-derived SSNs) was in the right ear (which had the recording microphone) or the left (contralateral) ear. By saving these additional averages, we were able to test the logical possibility that the amount of efferent activity differed in ears having, and not having, the targeted female voice—whether there was "ear suppression." (As will be seen, there was no systematic difference in the physiological responses from the ipsilateral and contralateral ears, within either subjects or conditions, an unexpected result.) For the auditory *inattention* condition, there were no incorrect trials (by definition), but the responses were kept separately for the ipsilateral and contralateral trials and for the two triplets, meaning that there were four averaged physiological measures per block of trials. (For the visual-attention study described below, the attention blocks were conceptually similar to the dichotic blocks in the auditory-attention study; thus, there were eight physiological averages at the end of each block. For the inattention conditions in the visual-attention study, there were four physiological averages at the end of each block.)

F. Analyzing nSFOAE responses

Following data collection, each averaged nSFOAE response (330 ms in duration) was analyzed offline by passing the averaged difference waveform through a succession of 10-ms rectangular analysis windows, beginning at the onset of the response, and advancing in 1-ms steps. At each step, the waveform segment was bandpass filtered between 3.8 and 4.2 kHz (the tonal signal was 4.0 kHz) using a sixth-order, digital elliptical filter. The rms voltage was computed for each 10-ms window and converted to dB SPL. Figure 2 shows an example of an nSFOAE response from subject
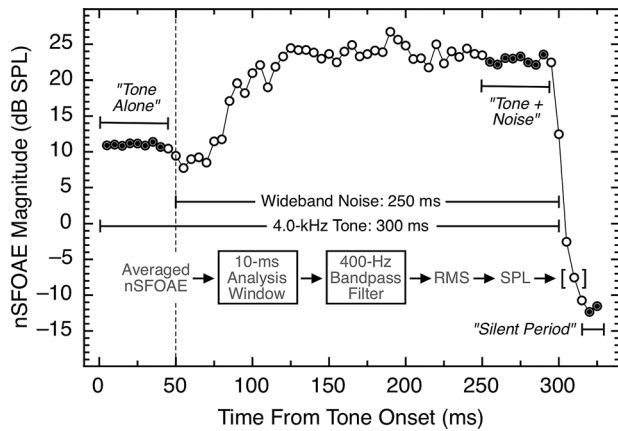
FIG. 2. An example of an nSFOAE response averaged from one block of trials of one condition. To summarize the data, tone-alone magnitudes were averaged from 5 to 40 ms, and tone-plus-noise magnitudes were averaged from 250 to 285 ms (thirty-five 10-ms epochs each; indicated by the filled circles). All nSFOAE responses, across all subjects and conditions, were analyzed in this same manner. The noise floor for these measurements was about −13.0 dB SPL. The data from the silent period already have been reported (Walsh *et al.*, 2014a,b).

L01, averaged from one block of trials. Time is shown along the abscissa, and nSFOAE magnitude is plotted on the ordinate. The analysis was performed in 1-ms steps, but only every fifth window is plotted in Fig. 2 to reduce data density.[2]

At the left in Fig. 2 is the nSFOAE response to the 50 ms of 4.0-kHz tone-alone, followed by the nSFOAE response to tone-plus-noise, in turn followed by the response during the 30-ms silent period. As can be seen, the response to tone-alone appears to be essentially instantaneous and essentially constant throughout the time course of the tone. By contrast, the response following the onset of the weak wideband noise was an initial dip, or "hesitation" of about 25 ms, followed by a rising dynamic response whose magnitude became essentially constant for the remainder of the tone-plus-noise stimulus. Walsh *et al.* (2010a) attributed the hesitation to mechanical suppression plus neural delay and attributed the rising dynamic phase to the onset of MOC activation (as described by Guinan *et al.*, 2003). The waveform morphology shown here was typical, but there were individual differences in nearly every aspect of the subjects' nSFOAE responses (the seemingly random fluctuations in the response are in fact attributable to the amplitude fluctuations in the particular sample of noise used). (The nSFOAE response shown in Fig. 2 comes from one block of trials, but for all analyses presented here, data were combined across several blocks, as is described below.)

Here we will emphasize the segment of the nSFOAE response seen during the tone-plus-noise portion of the eliciting stimulus rather than the segment of the response seen during the tone-alone portion. The reason is that tone-plus-noise is well known for its ability to trigger the MOC reflex (Guinan *et al.*, 2003; Walsh *et al.*, 2010a, 2014a); a rising, dynamic response having a time constant of about 100 to 200 ms is a common result (e.g., Backus and Guinan, 2006). Tone-alone is not as effective at triggering the MOC reflex, especially a tone as weak as 60 dB SPL (Guinan *et al.*,

2003). Evidence that our tone-alone stimulus was not triggering the MOC reflex is the absence of a rising, dynamic component to the nSFOAE response at tone onset; rather, the nSFOAE response to tone-alone had a rise time that matched that of the tonal stimulus for every subject. Accordingly, we believe that the nonlinearity underlying the tone-alone segment of our nSFOAE responses is primarily mechanical—no neural reflex is involved. Presumably the tone-plus-noise segment of our nSFOAE response involves some mechanical nonlinearity plus the additional nonlinearity associated with the MOC reflex. [We acknowledge that some of our averaged responses to tone-alone may contain some carryover of neural efferent effect from one test trial to the next (Backus and Guinan, 2006; Goodman and Keefe, 2006; Walsh *et al.*, 2010b, 2014a,b), which we call persistence.]

1. Culling and pooling of nSFOAE waveforms

As noted, each block of trials yielded four or eight averaged difference waveforms (nSFOAE responses), each of which was similar to the response shown in Fig. 2. Not all of these waveforms were judged to be acceptable for further analysis, however. Averaged responses having clearly atypical morphologies were excluded as follows: An nSFOAE response from a block of trials was eliminated from further analyses if (1) the asymptotic level of the tone-alone or tone-plus-noise portion of the response were within 3.0 dB of the noise floor as determined for that block of trials, or (2) if the average tone-plus-noise magnitude was more than 3.0 dB *smaller* than the average tone-alone magnitude. Using these criteria, only 18 of the 374 averaged nSFOAE responses collected across subjects, triplets, and conditions (∼5%) were culled.

Within blocks, difference waveforms obtained from individual triplets (trials) were summed point-for-point in time with other difference waveforms to yield final responses like those in Fig. 2. By averaging this way, random fluctuations were canceled out, and the weak nSFOAE response was allowed to emerge. *Across* blocks, however, the difference waveforms themselves were not summed; rather, the decibel values from corresponding 10-ms analysis windows were averaged. As described above, the nSFOAE responses from individual blocks were processed with a series of 10-ms windows (in 1-ms increments), each sample was filtered around the 4.0-kHz tone, and the rms amplitude of each window was calculated and transformed into decibels SPL. These decibel values then were averaged, window-by-window, across the nSFOAE responses obtained on different blocks of trials to determine the final overall averaged nSFOAE response for a condition for a subject. To summarize the responses resulting from this procedure, we averaged the nSFOAE magnitudes (in decibels) for the thirty-five 10-ms analysis windows beginning at 5 ms and for the thirty-five windows beginning at 250 ms (90% overlap for successive 10-ms windows). Those we call the *asymptotic levels* for the tone-alone and tone-plus-noise segments of the response, respectively (see Fig. 2).

J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh *et al.*: Efferents and attention   2743

2. Evaluating results

Numerous pairwise comparisons were of interest here, some within subjects and some across subjects. For each such comparison, an effect size (d) was calculated using the following equation:

$$d = (m_1 - m_2)/(((s^2_1{}^*(n_1 - 1)) + (s^2_2{}^*(n_2 - 1)))/(n_1 + n_2 - 2))^{1/2}, \quad (1)$$

where the numerator is the difference in the means (m) for the two conditions of interest, and the denominator is an estimate of the common standard deviation (s) for the two distributions. By convention, effect sizes larger than 0.2, 0.5, and 0.8 are considered small, medium, and large, respectively (Cohen, 1992).

For some comparisons, matched or unmatched t-tests also were used to evaluate possible differences between sets of data. For other comparisons, a bootstrapping procedure was implemented as a way of assessing the likelihood that the effect sizes obtained were attributable simply to chance. For each pair of conditions to be compared, all the data were pooled, a sample equal to the N of one of the conditions was drawn at random to simulate that condition, the remaining data were taken to simulate the second condition, an effect size was calculated for the two simulated conditions, and that value was compared to the effect size actually obtained in the study itself. This resampling process was repeated 20 000 times, and a tally was kept of how often the *absolute value* of the effect size in the resamples equaled or exceeded the effect size actually obtained in the study for that comparison. This tally, divided by 20 000, was taken as an estimate of the implied significance (p) of the actual outcome. (Our use of 20 000 resamples was a carry-over from our previous applications of this procedure, and although that number of resamples did well exceed the number of possible combinations for some of our comparisons, note that this over-sampling had no effect on our estimates of implied significance.) Also note that using the absolute values of the resampled effect sizes led to a relatively conservative ("two-tailed") assessment of implied significance. The resampling process was accomplished using custom-written LabVIEW software.

III. RESULTS: AUDITORY-ATTENTION STUDY

A. Behavioral tasks

The goal was to have a behavioral task that was neither too easy nor frustratingly difficult, and that goal was accomplished. All subjects performed above chance and below perfection at recognizing the middle five digits spoken by the female voice for both the dichotic and diotic auditory-attention conditions. Across the eight subjects, the average percent-correct performance was 86.4% and 86.6%, for the dichotic and diotic conditions, respectively. On average, each subject completed 4.7 blocks of 30 or more trials per condition. Five of the subjects performed better in the dichotic-listening condition, and three performed better in the diotic-listening condition. The two male subjects having

experience with early versions of the tasks (L01 and L05) found the diotic condition consistently, if only slightly, more difficult. Only one subject improved noticeably across the two test sessions (for both attention conditions), from about 71% correct to about 84% correct.

Whenever a subject responded within the 2.0-s response interval, a reaction time (RT) was calculated between the end of the time interval containing the seventh digit presentation and the key press. The values were saved separately for correct and incorrect responses except in the inattention condition, where all key presses were scored as "correct." Paired t-tests showed that RTs were significantly faster in the inattention condition than in the two attention conditions ($p \leq 0.025$, two-tailed and adjusted for multiple comparisons). This was expected because the inattention condition required no forced-choice decision about the content of the SSN stimuli heard during the trial. For both of the auditory-attention conditions, RTs were significantly slower for incorrect trials than for correct trials. For example, across the eight subjects in the dichotic auditory-attention condition, the mean RTs were 1520 and 1290 ms for incorrect and correct trials, respectively; unequal-variance $t(7) = 3.8$, $p < 0.01$. The results were similar for the diotic condition (and for the visual-attention conditions described below). There was no significant difference between the RTs for correct trials during the dichotic and diotic auditory-attention conditions. Thus, both the memory tests and the RT results confirm that all the subjects were actively engaged in the behavioral tasks.[1]

B. Cochlear responses during attention and inattention

For all subjects, there were differences in the magnitudes of the nSFOAE responses measured under conditions of attention and inattention. Typically, the differences were 2–3 dB. Before showing those results, we need to further explain the data analysis.

1. Dichotic condition: Ipsilateral and contralateral responses

When the auditory-attention study was designed, the dichotic condition was of particular interest because it provided the ability to test for "ear suppression." On approximately half of the trials in the dichotic condition, the attended female voice was in the right ear, the same ear as the microphone recording the nSFOAE responses. This was called the ipsilateral condition. On the other half of the trials, the female voice was in the left ear—called the contralateral condition. Given the neural circuitry of the OC efferent system (Brown, 2011), and various facts and speculations about the functioning of that system (Cooper and Guinan, 2006; Giard *et al.*, 1994; Guinan, 2006, 2010; Harkrider and Bowers, 2009; Jennings *et al.*, 2011; Kawase *et al.*, 1993; Kumar and Vanaja, 2004; Lukas, 1980; Walsh *et al.*, 2010b), we expected that there could be differences in nSFOAE magnitude on ipsilateral and contralateral trials, and accordingly, the data were saved separately for the two situations. However, for the majority of our subjects, the nSFOAE

2744   J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh *et al.*: Efferents and attention

responses for the ipsilateral and contralateral conditions were essentially the same (see below), and this was true for both triplets. The absence of ear suppression, while unexpected, allowed us to pool the data obtained from the ipsilateral and contralateral trials, and thereby increase the stability of our estimates for the dichotic condition. The pooling was accomplished not by summing the difference waveforms themselves, but by averaging the decibel values in the first 10-ms window of the averaged ipsilateral response (difference waveform) from each available block with that in the first window for the averaged contralateral response from each available block, and then repeating that process for each successive set of 10-ms windows in the two 330-ms responses (1-ms increments). In this way, 4–6 ipsilateral averaged responses were pooled with a corresponding number of contralateral responses.

To be precise about the similarities between the ipsilateral and contralateral nSFOAE responses: For each subject, statistical comparisons were made between the asymptotic values estimated for the tone-alone segments for triplet 1, the tone-plus-noise segments for triplet 1, and the same for triplet 2. Only two of those 32 statistical comparisons even approached marginal significance: (1) For all four blocks that L03 completed for the dichotic condition, the asymptotic magnitude for tone-plus-noise on triplet 2 was greater for the ipsilateral than the contralateral trials; the average difference was 1.6 dB. (2) For all six blocks that L05 completed for the dichotic condition, the asymptotic magnitude for tone-alone on triplet 2 was greater for the ipsilateral than the contralateral trials; the average difference was 2.8 dB. For L03, the p-value for the comparison for tone-plus-noise for triplet 2 was 0.052, and for L05, the p-value for the comparison for tone-alone for triplet 2 was 0.047 (matched t-tests). After Bonferroni correction for the 32 comparisons, these two differences did not achieve statistical significance, and the 30 other comparisons all had smaller differences than these. Accordingly, we felt it was appropriate to pool the ipsilateral and contralateral data for all subjects.[1]

For the inattention condition in the auditory-attention study, the physiological averages were averaged and stored separately depending upon whether the SSNs derived from the female voice were in the ipsilateral or contralateral ear, even though that manipulation was of no consequence for the subject's behavioral response. To obtain final averages, those ipsilateral and contralateral responses were combined as described above for the dichotic-attention condition. Matched t-tests showed no significant differences between the "ipsilateral" and "contralateral" responses in the inattention condition, and this was true for all subjects. This null outcome served as an important assessment of our procedures. Compared to the dichotic condition, where ipsilateral and contralateral differences seemed plausible, the same division of data in the inattention condition showed that such differences were unlikely to occur due to chance. (For the visual-attention study described below, the "ipsilateral" and "contralateral" averages also were combined to yield the final averaged response for each block of trials. No significant statistical differences were observed; as in the other conditions just described, the "ipsilateral" and

"contralateral" responses were essentially the same within all subjects.) Accordingly, we conclude that the physiological mechanisms operating in our study were essentially equally effective in the two ears.

2. Inattention and attention conditions compared

Representative examples of the nSFOAE responses obtained in this study are shown in Fig. 3. For this subject, the responses during most of the tone-alone and tone-plus-noise periods were larger during inattention than during attention. [This difference also was observed in the 30-ms silent period, as was reported and discussed by Walsh *et al.* (2014a,b).] For this subject, similar differences in responses for the inattention and dichotic conditions were obtained for triplet 2, and these differences were present as well when the inattention responses for both triplets were compared to those of the diotic-attention condition.

As noted above, the asymptotic nSFOAE responses for tone-alone and tone-plus-noise portions of the response were summarized by averaging the decibel values across thirty-five 10-ms windows beginning at 5 and at 250 ms, respectively (1-ms increments) from waveforms like those shown in Fig. 3. The results for the tone-plus-noise portion of the response are shown in Fig. 4. Note that for the five subjects grouped at the left in Fig. 4, the asymptotic values for tone-plus-noise were *greater* in the inattention condition than in either of the auditory-attention conditions, and for the three subjects at the right in Fig. 4, the asymptotic responses were *smaller* in the inattention condition than in either attention condition. The *direction* of the difference always was the same for a given subject, across multiple blocks of trials both within and across test sessions; it just was not the same across subjects. We ask the reader to ignore these individual differences in the direction of effect for the moment; a
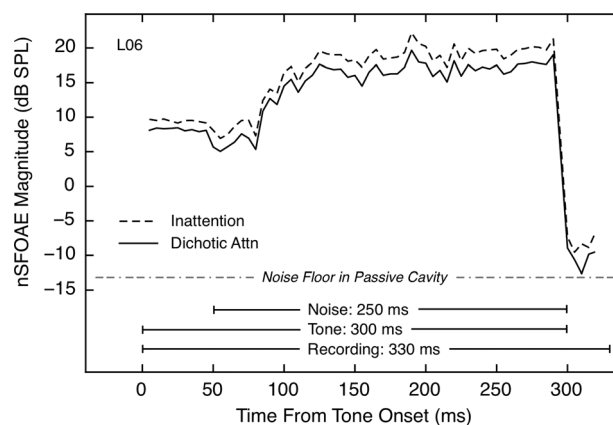

FIG. 3. The nSFOAE responses for the full 330-ms measurement period for one subject for both the auditory inattention condition and the dichotic attention condition. The first 50 ms shows the response to tone-alone, the next 250 ms shows the response to tone-plus-noise, and the final 30 ms shows the response after the stimuli terminated. The dichotic response represents the mean of the nSFOAE averages obtained when the attended voice was either in the ear ipsilateral to the microphone or in the contralateral ear. The fine structure of the response during tone-plus-noise is not random fluctuation, but rather, it is attributable to the specific sample of noise used. Each response included only trials having correct behavioral responses and was averaged across at least four 30-trial blocks.
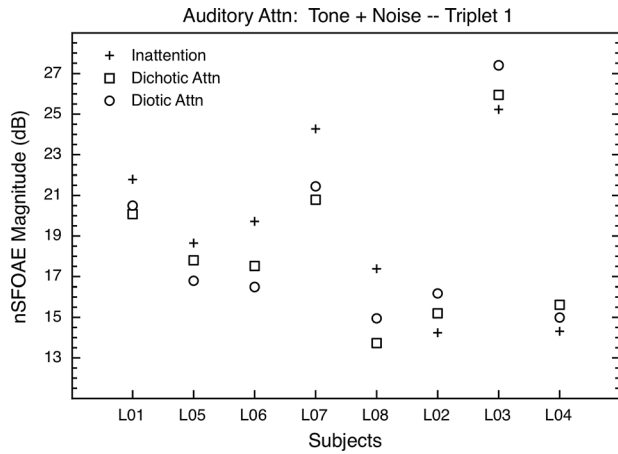
FIG. 4. Auditory-attention study: The asymptotic nSFOAE responses to tone-plus-noise for triplet 1 for all subjects in the auditory study. The values shown began as means across thirty-five 10-ms analysis windows beginning at 250 ms into the 300-ms nSFOAE-eliciting stimulus within one block of trials (i.e., at response asymptote); those individual means then were averaged across at least four such blocks. The differences between inattention and attention were in one direction for the five subjects at the left of the figure, and in the other direction for the remaining three subjects. Standard errors of the mean were 0.73 dB, on average. The level of the noise floor of our measurement system at 4.0 kHz was about −13.0 dB SPL (see footnote 2).

plausible explanation is provided in Sec. VII A below. For now, it is simply the existence of differences between attention and inattention that is of interest.

The differences between inattention and attention in Fig. 4 are in the range of 2–3 dB, and thus are modest in size. However, these differences are substantially larger than many attentional differences reported using tone-evoked or click-evoked OAEs (Puel *et al.*, 1988; Froehlich *et al.*, 1990; Froehlich *et al.*, 1993; Meric and Collet, 1992, 1994; Giard *et al.*, 1994; Ferber-Viart *et al.*, 1995; Maison *et al.*, 2001; Harkrider and Bowers, 2009). Also, the decibel differences do not fully capture the magnitudes of our attentional differences within subjects. As evidence, we calculated effect sizes for the differences between the inattention and attention conditions within individual subjects. The results are shown in Table I. Effect sizes are shown separately for the dichotic and diotic conditions, each compared to the inattention condition. The top and bottom halves of the table contain the effect sizes for tone-plus-noise and tone-alone, respectively, and the left and right halves of the table pertain to the two triplets.

As Table I reveals, the differences between attention and inattention were far more substantial for most subjects than the 2–3 dB differences of Fig. 4 suggest. Contributing to these large effect sizes is the low variability within subjects across blocks of trials. The differences in sign in Table I reflect the individual differences in direction of effect seen in Fig. 4. In accord with our intention to temporarily ignore the differences in direction of effect, the mean effect sizes shown across subjects are means of the *absolute values* of the individual effect sizes. Table I reveals that the differences between inattention and attention generally were smaller for triplet 2 than for triplet 1, and generally smaller for tone-alone than for tone-plus-noise.

TABLE I. Auditory-attention study: Effect sizes for the differences between the inattention condition and both attention conditions, shown separately for both triplets.

Subject Number	Triplet 1		Triplet 2	
	Inattention minus Dichotic	Inattention minus Diotic	Inattention minus Dichotic	Inattention minus Diotic
Tone + Noise				
L01	0.79	0.70	0.48	−0.17
L02	−0.40	−0.89	−0.31	−1.16
L03	−0.39	−1.10	−1.62[a]	−1.49[b]
L04	−0.89	−0.45	−0.10	0.15
L05	0.65	1.27[b]	−0.30	−0.28
L06	1.47[b]	1.83[b]	0.34	0.23
L07	1.51[b]	1.39	0.89	0.51
L08	1.73[a]	1.07[b]	0.13	0.14
Mean (absolute values)	**0.98**	**1.09**	**0.52**	**0.52**
Tone Alone				
L01	0.32	0.22	0.38	0.02
L02	−0.48	−1.10	0.37	−0.55
L03	−0.66	−1.00	−0.34	0.12
L04	0.00	0.19	−0.70	−0.97
L05	−1.67[a]	−0.87	0.27	−0.04
L06	0.68	0.86	−0.21	−0.24
L07	0.95	0.47	0.93	0.17
L08	1.38[b]	1.09	0.35	0.17
Mean (absolute values)	**0.77**	**0.73**	**0.44**	**0.29**

[a]$p \leq 0.01$ (Implied statistical significance computed using a bootstrapping procedure.)
[b]$p \leq 0.05$ (Implied statistical significance computed using a bootstrapping procedure.)

IV. VISUAL-ATTENTION STUDY

The second study was highly similar to the first, the primary difference being that the attended stimuli were presented visually. Seven of the eight subjects from the auditory-attention study participated in the visual-attention study. Again, there were at least two 2-h test sessions conducted for every subject, and, for each condition, data were collected for at least four blocks of trials consisting of at least 30 trials each. As in the auditory-attention study, Etymotic earphones were fitted in both ear canals, and an Etymotic microphone was in the right ear only. The collection and analysis of nSFOAE responses was the same as in the auditory study.

A. Visual-attention task

The subject again was presented with two strings of seven digits on every trial, was asked to attend to one of them and, at the end of each trial, was required to recognize the middle five digits of the attended string. The two strings of digits were presented visually, side-by-side, for 500 ms each, with one sequence inside a pink box (the attended string) and one inside a blue box. The pink box was on the left side of the display on half of the trials at random, and remained on that side throughout a trial. Interleaved with the seven pairs of visual digits were the six, diotic nSFOAE-eliciting sounds, each lasting 330 ms. In some conditions, sequences of spoken

digits (attention condition) or SSNs (inattention condition) were presented as additional distracters; when present, they were dichotic and simultaneous with the visual digits. The durations of the digits and eliciting sounds were the same as for the auditory-attention study. The trial-timing sequence, including the 2000-ms response interval and the 200-ms feedback interval, is shown in Fig. 5. For the inattention conditions, the subject was expected only to press a response key after the final pair of pink and blue boxes was presented. RTs were collected for all behavioral responses. The pink and blue sequences of visual digits (and the dichotic pairs of spoken digits or SSNs when they were presented) all were selected independently.

There were two attention conditions and two inattention conditions in the visual-attention study. For both *inattention* conditions, no visual digits were presented in the pink and blue boxes, and all the subject was required to do was press a response key immediately after the final pair of boxes was presented. For one attention and one inattention condition, the only auditory stimuli were the nSFOAE elicitors; neither auditory digits nor SSNs were presented. For the second *attention* condition, auditory digits were presented simultaneous with the visual digits. These presentations were dichotic—the female voice was in one ear and the male voice in the other ear at random—and they spoke strings of digits of SSNs that were independent of each other and of the visual strings being presented. For the second *inattention* condition, SSNs were presented in place of the auditory digits.

Each subject adjusted the viewing distance to the monitor for his or her comfort. Those distances were not routinely measured, but the visual angles subtended by the digits unquestionably did vary across subjects. Under reasonable assumptions about viewing distance, the visual angle subtended was about 1.2–1.6 degrees for the height of the digits,

about 2.4–3.2 degrees for the sides of the pink and blue boxes, and about 3.6–4.8 degrees for the lateral separation between the digits. Thus, the visual digits were large and well separated.

V. RESULTS: VISUAL-ATTENTION STUDY

A. Behavioral tasks

For the visual-attention study, just as for the auditory-attention study, performance on the digit-identification task during the attention conditions and the RTs during the inattention conditions revealed that all subjects were actively engaged in the behavioral tasks. Across the seven subjects, the percentage of correct decisions about the target digits was similar to that in the auditory-attention study; for the visual-attention condition without speech sounds, the average percent-correct performance was 92.0%, and for the visual-attention condition with speech sounds performance was 90.1% correct on average. Also similar to the auditory-attention study, reaction times were significantly faster for the visual-*inattention* conditions than for the visual-attention conditions ($p \leq 0.05$, two-tailed), and RTs were significantly faster for correct versus incorrect trials ($p \leq 0.05$, two-tailed).[1]

B. Cochlear responses during attention and inattention

In the visual-attention study, just as in the auditory-attention study, there were differences in the magnitudes of the nSFOAE responses for the inattention and attention conditions, and this was true whether or not speech or SSNs were present as additional distracter stimuli. The asymptotic
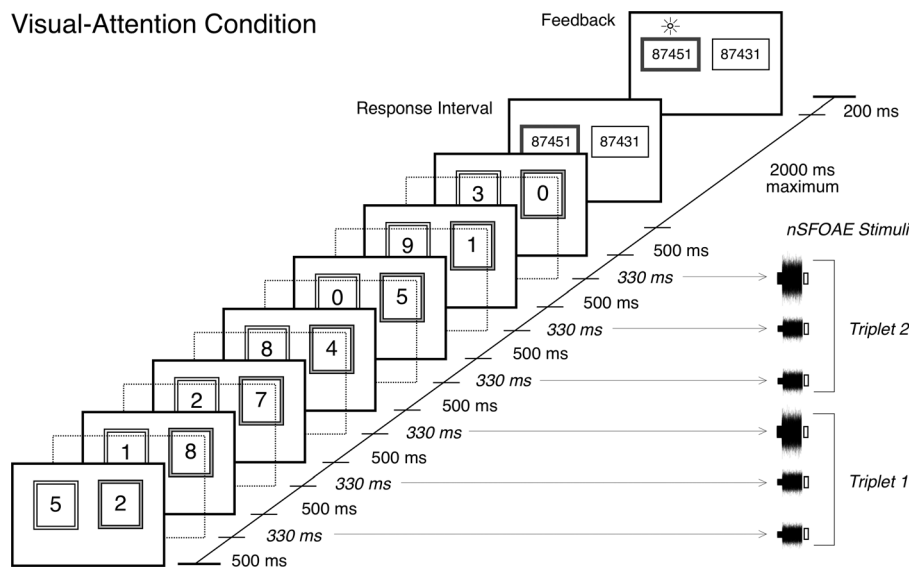

FIG. 5. Trial-timing sequence for the visual-attention study. Sequences of seven pairs of digits were presented simultaneously for 500 ms each, one sequence inside a pink box (gray in the figure) and one inside an adjacent blue box. Interleaved with the seven pairs of visual digits were six diotic presentations of the nSFOAE-eliciting stimuli, each being 330 ms in duration. Following the last pair of visual digits was a response interval of 2000 ms, during which the subject pressed one of two response keys to indicate which of two sequences of five digits contained the middle five digits of the digit sequence in the pink box. Immediately following was a feedback interval of 200 ms, during which a small symbol was displayed over the correct five-digit choice. In the inattention conditions, the subject pressed a response key immediately after the last nSFOAE stimulus was presented (no digits were displayed). In some conditions, dichotic speech or SSNs were presented simultaneously with the visual digits. [Reprinted with permission from Walsh *et al.* (2014b).]
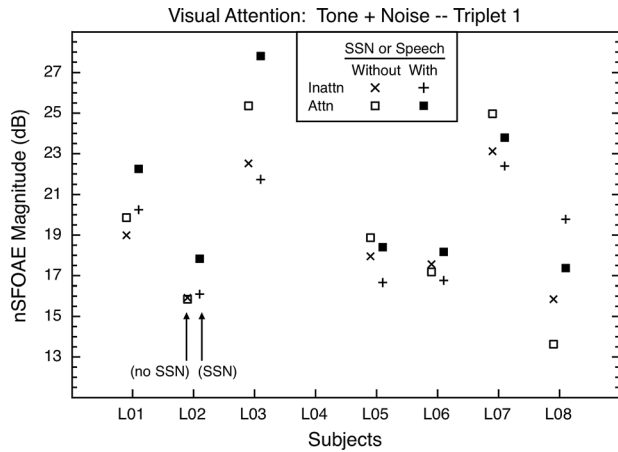
FIG. 6. Visual-attention study: The asymptotic nSFOAE responses to tone-plus-noise for triplet 1 for all subjects in the second study. One set of inattention and attention conditions (With) included spoken digits during the attention blocks and SSNs during the inattention blocks; another set included no speech nor SSN stimuli (Without). The values shown began as means across thirty-five 10-ms analysis windows beginning at 250 ms into the 300-ms nSFOAE-eliciting stimulus within one block of trials; those individual means then were averaged across at least four such blocks. As in the auditory-attention study, the differences between inattention and attention were in different directions for the different subjects. Within subjects, standard errors of the mean were about 0.78 dB, on average.

responses to tone-plus-noise for triplet 1 are shown in Fig. 6. The data from triplet 2 were similar.

1. Inattention and attention conditions compared

The pair of inattention and attention conditions having no spoken digits or SSNs present was the simplest test of whether the cochlea was affected during visual attention. Examination of Fig. 6 reveals that there *were* attentional differences between these conditions, although they generally were smaller than the attentional differences for the pair of conditions in which spoken digits and SSNs also were present as distracters. Again, the decibel differences between inattention and attention underestimated the actual attentional differences; see the within-subject effect sizes in Table II. The left and right halves of the table contain the effect sizes for the two triplets, and the top and bottom halves pertain to tone-plus-noise and tone-alone, respectively. Effect sizes are shown separately for the conditions that did, and did not, contain speech or SSNs as auditory distracters during visual attention, and as in Table I, absolute values were used when calculating the means of the effect sizes across subjects. Note in Fig. 6 and Table II that there again were individual differences in whether the attention or inattention condition produced the stronger nSFOAE, just as was true for the auditory-attention study. Furthermore, individual subjects often did not exhibit the same directionality of effect in the two studies.

Table II reveals that attending to visual stimuli did produce differences in the cochlea during the tone-plus-noise portion of the eliciting stimulus. In accord with intuition, additional distracting stimuli did matter for most subjects. The attentional differences generally were larger when spoken digits or SSNs were present than when they were not.

TABLE II. Visual-attention study: Effect sizes for the differences between the inattention and attention conditions, shown separately for both speech/SSN conditions and both triplets.

	Triplet 1		Triplet 2	
Subject Number	Inattention minus Attention Without SSNs[a]	Inattention minus Attention With SSNs	Inattention minus Attention Without SSNs	Inattention minus Attention With SSNs
Tone + Noise				
L01	−0.33	−0.92	−1.14	1.80[b]
L02	0.02	−0.84	0.02	0.42
L03	−0.51	−1.92[b]	−1.15[c]	2.17[b]
L04	—	—	—	—
L05	−0.78	−0.88	−0.38	1.12[c]
L06	0.23	−1.63[b]	−0.05	0.38
L07	−0.94	−0.95	−0.62	0.31
L08	1.12[c]	1.30[c]	−0.50	0.25
Mean (absolute values)	**0.56**	**1.21**	**0.55**	**0.92**
Tone Alone				
L01	1.51[c]	−0.79	0.48	0.52
L02	1.04	−0.60	0.18	0.20
L03	−0.75	−0.44	−0.47	−0.30
L04	—	—	—	—
L05	−0.32	−0.77	−0.44	−0.18
L06	0.43	−0.39	0.43	−0.41
L07	−0.49	−0.74	1.16[c]	−0.21
L08	1.05[c]	0.85	−0.17	−1.14[c]
Mean (absolute values)	**0.80**	**0.65**	**0.48**	**0.42**

[a]SSN: speech-shaped noise stimuli, presented dichotically, and used as auditory distracters. When SSNs were used in the inattention condition (WITH), dichotic spoken digits were used in the corresponding attention condition.
[b]$p \leq 0.01$ (Implied statistical significance computed using a bootstrapping procedure.)
[c]$p \leq 0.05$ (Implied statistical significance computed using a bootstrapping procedure.)

For 24 of the 28 entries for tone-plus-noise in Table II, the effect size was greater in absolute value when speech or SSNs were present than when not. Moreover, when spoken digits or SSNs were present in the visual-attention study, the effect sizes were generally quite similar to those in the auditory-attention study, and that was true for both tone-plus-noise and tone-alone (compare Tables I and II). These similarities constitute a form of replication for the auditory-attention study, although the subjects were the same.

VI. ADDITIONAL FINDINGS AND COMMENTS

A. Time delay of the SFOAE

Although we have no direct measures of the time delays of the SFOAE itself, we did calculate FFTs for every 10-ms analysis window, and each of those calculations provided a phase value for the 4.0-kHz component of the nSFOAE. Figure 7 compares the level and phase data for two representative subjects, one showing a phase change of about 100 degrees from tone-alone to tone-plus-noise, and the other showing a phase change of less than 25 degrees. Three of the eight subjects showed the larger phase-change pattern seen
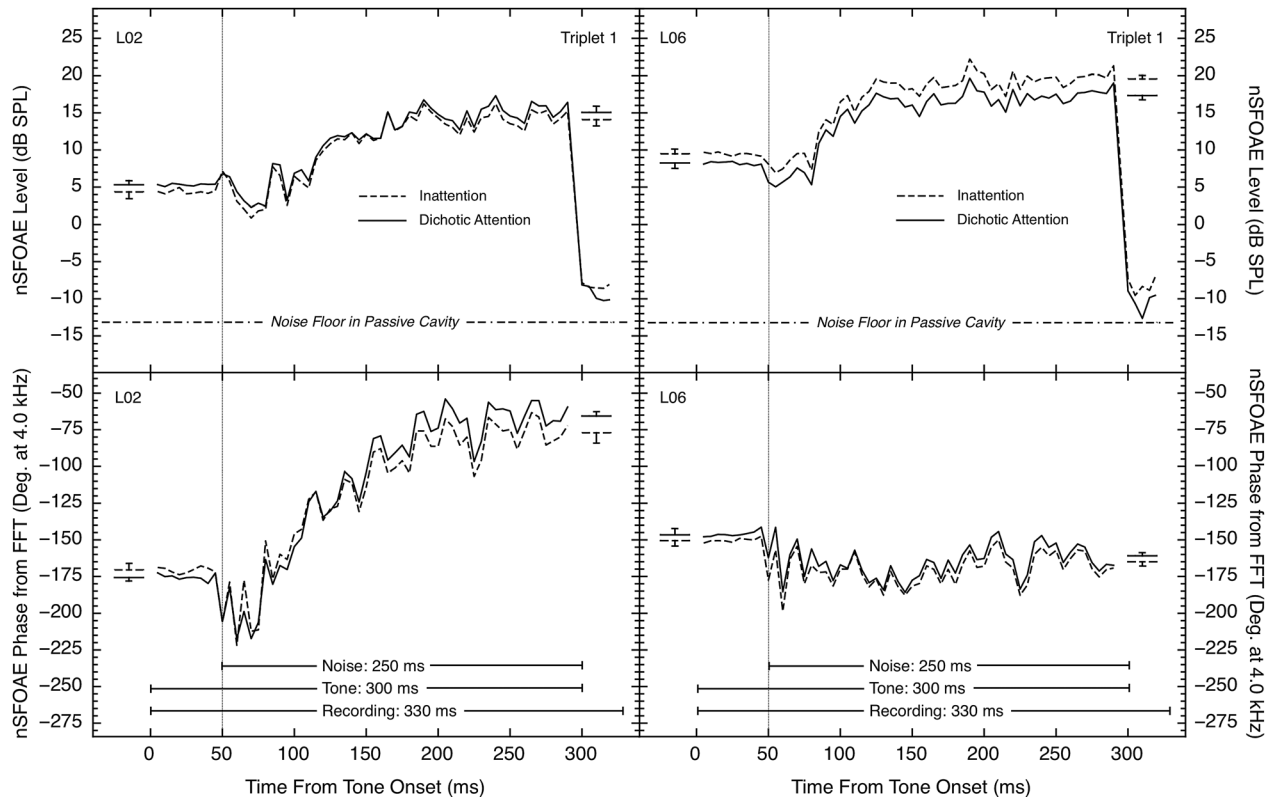
FIG. 7. Comparison of (top) level and (bottom) phase of the nSFOAE response at 4.0 kHz for triplet 1 for two subjects from the auditory-attention study. The timing lines in the bottom panels mark the tone-alone and tone-plus-noise portions of the response. Level measurements were obtained for each successive 10-ms analysis window using a filter centered at 4.0 kHz and having a bandwidth equal to 10% of its center frequency; the phase values were obtained by performing FFTs on each 10-ms analysis window. At the left and right within each panel, plus or minus one standard error is shown for the estimates of the asymptotic tone-alone and the asymptotic tone-plus-noise responses, respectively; standard errors were calculated across the multiple blocks that were averaged.

on the left side of Fig. 7, and five showed relatively small phase changes. Each subject showed similar phase patterns in both the auditory and visual experiments. As can be seen, for some subjects, phase changed in a manner paralleling the changes in level (on the left in Fig. 7), but for other subjects phase did not show the same rising, dynamic pattern as level (on the right in Fig. 7).

Note that these measures of phase are not simple because they are the phase of a "waveform" that results from subtracting the response to the single two-earphone presentation from the sum of the responses to the two single-earphone presentations. Also, the initial values (at time 0 ms) are necessarily constrained to be within a single period of the 4.0-kHz tone, meaning that a large difference in the time delay of the SFOAE (say, between the inattention and attention conditions) logically could be underestimated by one or more periods. Throughout the 300-ms interval of the recording, when the 4.0-kHz tone was present, phase values did remain within one period. Large, rapid phase changes were never observed in our succession of 10-ms analysis windows after the onset of the wideband noise. Larger rapid phase changes were occasionally seen during the initial 25 ms following noise onset, where the nSFOAE level dipped briefly, making phase measurements less reliable. As was the case for the level functions, the fine grain of the phase functions was very similar in the attention and inattention conditions, again suggesting that these fluctuations were due to the

characteristics of the wideband noise stimulus, not random variability in the data.

For consistency with the magnitude data, the phase values for the thirty-five 10-ms analysis windows beginning at 5 ms and at 250 ms were averaged to produce estimates of the asymptotic phase values for the tone-alone and tone-plus-noise portions of the response. This was done for the nSFOAE response from each block of trials for each subject, and the resulting values were averaged across the repeated measures of that condition for that subject. Those asymptotic phase values are shown in Table III. The tone-plus-noise data are shown at the top of the table, the tone-alone data at the bottom of the table, and the data for triplets 1 and 2 are shown at left and right, respectively. Means of the absolute values of phase are shown below the data for individual subjects.

For the purpose of comparing the differences in phase values between the inattention and attention conditions, effect sizes were calculated, just as was done for the level data (Tables I and II); those calculations are shown in Table IV. The format of the table is the same as for Table III. The effect sizes in Table IV are, on average, medium or large in value. The reason is that, just as for the level measure, the asymptotic phase values exhibited quite small variability across blocks of trials. The data in Table IV are from the auditory-attention study, but the data for the visual-attention study were quite similar.

TABLE III. Auditory-attention study: Asymptotic phase values at 4.0 kHz from FFTs for inattention and attention conditions, shown separately for tone-plus-noise (top) and tone-alone (bottom), and for both triplets.

Subject Number	Triplet 1			Triplet 2		
	Inattention	Dichotic	Diotic	Inattention	Dichotic	Diotic
Tone + Noise						
L01	−161.33	−152.37	−153.67	−159.19	−150.94	−150.75
L02	−77.09	−65.66	−63.28	−74.42	−64.34	−60.69
L03	−39.19	−38.53	−43.96	−37.93	−44.08	−44.89
L04	−36.46	−46.77	−42.33	−41.36	−49.02	−46.24
L05	125.12	117.46	122.26	121.63	119.31	123.61
L06	−165.02	−160.93	−162.94	−160.96	−160.99	−160.37
L07	−140.06	−118.32	−136.48	−145.01	−130.67	−156.26
L08	−145.64	−133.18	−126.63	−136.07	−138.22	−133.81
Tone Alone						
L01	−194.15	−179.30	−181.17	−221.86	−202.62	−210.12
L02	−170.55	−175.76	−175.45	−203.93	−201.18	−211.77
L03	26.97	24.40	17.10	38.97	43.47	43.65
L04	−88.88	−92.99	−93.08	−151.07	−144.21	−154.29
L05	117.39	123.75	127.90	24.04	10.37	5.13
L06	−150.55	−146.58	−148.62	−86.40	−129.23	−123.28
L07	−147.57	−132.34	−135.31	−25.72	−33.12	−27.33
L08	−149.61	−141.69	−134.76	−132.15	−144.47	−142.52

B. Frequency extent of attentional differences

All nSFOAE magnitude data presented to this point were obtained with the analysis filter centered at 4.0 kHz, the frequency of the tone in our MOC-eliciting stimulus. By

TABLE IV. Auditory-attention study: Effect sizes for the phase differences between the auditory inattention condition and both auditory-attention conditions, shown separately for both triplets.

Subject Number	Triplet 1		Triplet 2	
	Inattention minus Dichotic	Inattention minus Diotic	Inattention minus Dichotic	Inattention minus Diotic
Tone + Noise				
L01	−0.97	−1.05	−0.92	−1.06
L02	−0.77	−0.85	−0.68	−0.80
L03	−0.09	0.58	0.78	0.80
L04	0.89	0.42	0.87	0.50
L05	0.62	0.21	0.19	−0.15
L06	−0.76	−0.33	0.01	−0.11
L07	−0.77	−0.11	−0.51	0.38
L08	−1.13	−1.25	0.18	−0.15
Mean (absolute values)	0.75	0.60	0.52	0.49
Tone Alone				
L01	−1.42	−1.92	−0.85	−0.55
L02	0.51	0.42	0.33	0.47
L03	0.50	1.79	−1.28	−0.97
L04	0.35	0.38	−0.39	0.24
L05	−0.54	−0.88	0.48	0.62
L06	−0.37	−0.18	0.86	0.66
L07	−0.78	−0.76	0.30	0.30
L08	−0.83	−1.06	0.56	0.43
Mean (absolute values)	0.66	0.92	0.63	0.53

selecting different center frequencies of the analysis filter, we were able to examine the magnitude of the attentional difference across the spectrum (for consistency, we kept the bandwidth of the filter at 10% of the center frequency and then corrected the data to compensate for the different rise times of those analysis filters; see Walsh *et al.*, 2014a,b).[2] Once the analysis filter was off the frequency of the 4.0-kHz tone, the measurements were of the nSFOAEs to the relevant frequency components of the weak wideband noise. The data were analyzed separately for the auditory- and visual-attention studies, for the inattention and attention conditions, and for the tone-alone, tone-plus-noise, and silent periods of the acoustic stimulus. As was true for the 4.0-kHz data above, the directionality of the difference between inattention and attention could differ across frequency within a subject, it could differ across subjects at individual frequencies, and the patterns were not the same for all subjects. This fact again made simple means of the inattention/attention difference inappropriate, so again the absolute values of the differences between inattention and attention were used.

In the top panel of Fig. 8 are the mean data from only the relevant *inattention* conditions for the tone-plus-noise, tone-alone, and silent-period portions of the nSFOAE response as the center frequency of the analysis filter was varied. As can be seen, these data confirm two expectations: (1) the magnitudes of the nSFOAEs for tone-plus-noise were considerably stronger than those for both tone-alone and the silent period; and (2) the magnitudes of the nSFOAEs at, and close to, 4.0 kHz were considerably stronger than those at more distant frequencies (where the acoustic stimulus was much weaker). Both outcomes are in accord with past experience with nSFOAE behavior as level was varied (see Walsh *et al.*, 2010a,b). Also evident from the top panel of Fig. 8 is that the responses from the inattention conditions of the auditory- and visual-attention studies were remarkably
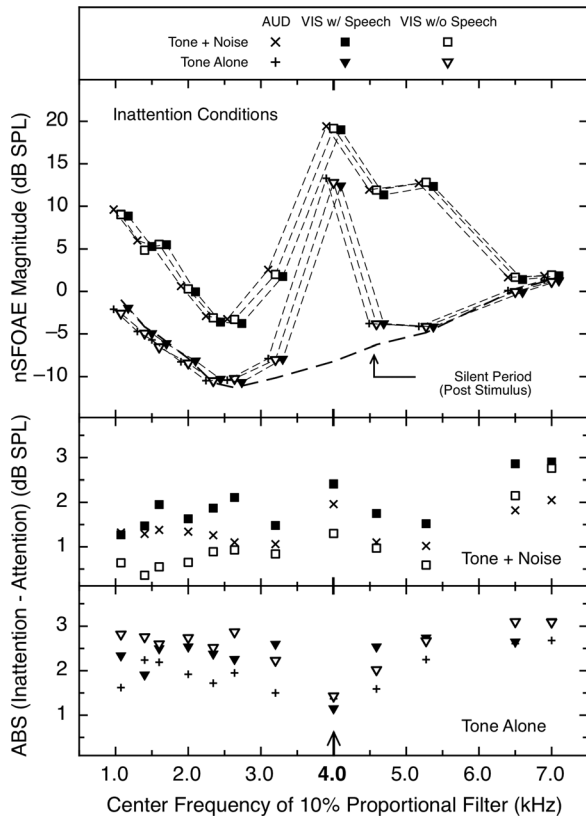
FIG. 8. Attentional effects across the spectrum. (top) Average magnitude of the asymptotic nSFOAE response for the *inattention* conditions in the auditory- and visual-attention studies. Data are shown for the tone-alone, tone-plus-noise, and silent periods of the eliciting stimuli. The nSFOAE-eliciting stimulus was a moderate 4.0-kHz tone and a weak wideband noise. The values shown were obtained by varying the center frequency of the analysis filter and repeating the moving-window analysis. All values have been corrected for the rise times of those different filters (see Walsh *et al.*, 2014a,b), and for clarity, entries have been slightly displaced laterally. The noise floor of the measurement system lay several decibels below the values from the silent period. (Middle) and (bottom) Across-subject means of the absolute values (ABS) of the *differences* between the inattention and attention conditions for the tone-plus-noise and tone-alone portions of the nSFOAE responses, respectively. For the auditory-attention study, the average standard error of the differences of the means (Inattention − Attention) across subjects and frequencies was 0.33 dB; for the visual-attention study, it was 0.37. For the auditory-attention conditions there were eight subjects; for the two visual-attention conditions only seven of those subjects completed the study.

similar in magnitude and pattern across frequency (presumably, in part, because the same sample of wideband noise was used for all subjects in both studies). (Supplemental analyses revealed that the flat segment in the vicinity of 5.0 kHz in the tone-plus-noise data exists even when the sampling of frequency values is more dense.)

The two bottom panels of Fig. 8 contain the means of the *absolute values* of the within-subject *differences* between inattention and attention, with the tone-plus-noise differences in the middle panel and the tone-alone differences in the bottom panel. At 4.0 kHz, the attentional differences were larger for tone-plus-noise than for tone-alone, but the reverse was true at a number of other analysis frequencies. As noted above, we believe that the response to tone-alone originates in a mechanical nonlinearity (e.g., that response shows no rise time), whereas the response to tone-plus-noise involves

both that mechanical nonlinearity plus a neural one (associated with the MOC reflex). For the present tasks, at least, both of those mechanisms were affected by the attention manipulation, and both were affected across quite a wide band of frequencies. (In Sec. VII C, the possible role of short-term persistence of MOC efferent activity in these tone-alone results is discussed.) Note that the two highest frequencies shown in Fig. 8 lay above the high-frequency cutoff of our wideband noise, and for all subjects, the response magnitude was greater for the inattention conditions than for the attention conditions at those frequencies, just as was true for the silent period (Walsh *et al.*, 2014a,b). Thus, for our task and stimuli at least, efferent effects were quite wideband.

VII. SUMMARY AND DISCUSSION

In both studies, the nSFOAE responses obtained during the tone-plus-noise portion of the eliciting stimulus were different when subjects were attending, or not attending, to the auditory or visual digits. The differences between attention and inattention conditions were only about 2–3 dB, but within-subject effect sizes for those differences were moderate to large, and this was true for both the auditory and visual-attention studies. Because our nSFOAE-eliciting stimuli were interleaved with the auditory or visual digits, our cochlear measures were obtained *during* the two attentional states. Furthermore, our nSFOAE measures were perstimulatory—made *during* the presentation of the OAE-evoking stimulus—rather than after its offset, as is the case for tone-evoked and click-evoked OAEs. This may have provided an advantage over some previous procedures (e.g., de Boer and Thornton, 2007; Ferber-Viart *et al.*, 1995; Giard *et al.*, 1994; Harkrider and Bowers, 2009; Meric and Collet, 1994; Puel *et al.*, 1988).

Note that "inattention" was a relative matter here. In our inattention conditions, the subjects *did* have to attend to the trial-timing sequence in order to make the required motor response during the response interval, and their RTs were fast. These subjects were not navel-gazing; they simply did not have to attend to any digits, nor perform a forced-choice task at the end of each trial. Accordingly, the attention/inattention differences observed here likely are underestimates of what would be obtained with a less-involving inattention task (see Fritz *et al.*, 2007, for a discussion of the issues surrounding measuring and modulating auditory and visual attention).

As noted, the nSFOAE responses to tone-plus-noise obtained in these studies (Figs. 4 and 6) typically included a rising, dynamic segment that began about 25 ms after noise onset, and took about 100–150 ms to asymptote. This is commonly interpreted as a defining characteristic of the reflexive activation of the MOC efferent system (Backus and Guinan, 2006; Guinan *et al.*, 2003; Walsh *et al.*, 2010a,b). The response to tone-alone, on the other hand, always was quite rapid, appearing to be limited only by the rise-time of the tonal eliciting stimulus (also see Walsh *et al.*, 2010a,b). This suggests that, although pure tones sometimes *can* activate

J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh *et al.*: Efferents and attention   2751

the MOC efferent system (Guinan *et al.*, 2003), the 4.0-kHz tone in the current studies was not doing so.

A. Different directions of effect

In both studies, some subjects showed greater magnitudes of nSFOAE response for the inattention condition than for the attention condition, and some subjects showed the reverse. Also, the direction of effect shown by a subject sometimes was different in the two studies. Within subjects and conditions, however, the directionality was highly consistent across blocks of trials and across sessions, strong evidence that this was not simply random variation (see the large within-subject effect sizes in Tables I and II). One interpretation is that these different directions of effect are accurate indications that different brain processes exist in different subjects; that is, some people really do have greater MOC activation during attention than inattention and other people have the reverse. In our opinion, this interpretation is highly unlikely to be correct. We have no hard evidence to offer, but we believe that a hindbrain circuit like the MOC reflex evolved to solve some fundamental problem faced by the auditory nervous system, a problem that is common to all members of each species having the reflex (see Sec, VII G below). If the problem is common to all individuals of a species, then it is logical that the solution provided by the reflex also should be common to all individuals. For us at least, logic, intuition, and parsimony suggest that a highly conserved network like the MOC reflex works in the same direction for all subjects, and that some other factor is operating to produce the individual differences in direction of effect seen in our measures. Our belief is bolstered by the existence of a plausible candidate for this "other factor."

We believe that the different directions of effect across subjects are attributable to individual differences in the time delay with which the reflected emission propagates back to the ear canal. For simplicity, let us first consider the SFOAE before moving to the nSFOAE. Various lines of evidence suggest that activation of the MOC system makes the cochlear partition stiffer (e.g., Brown *et al.*, 1993). Let us assume (plausibly) that, for *all* of the subjects in our two studies, this increase in stiffness was greater during attention than during inattention (and that this was true over most of the basilar membrane, see Fig. 7). Assume further that the greater stiffness during attention led to a slight decrease (say) in the time delay with which the SFOAE reached the ear canal compared to the inattention condition.

Our measurements were not of the SFOAE, but phase interactions are relevant for the nSFOAE. Whatever the phase of the SFOAE during the first presentation of each triplet, that phase value should be the same for the second presentation because the only difference was the earphone delivering the stimulus. For the third presentation of each triplet, however, the overall level of the stimulus was about 6 dB greater, meaning that the MOC-induced stiffness of the basilar membrane should be greater than for either of the first two presentations of the triplet. Also note that the *level* of the SFOAE likely grows compressively across that 6-dB range (Schairer *et al.*, 2003), meaning that the SFOAE in the

ear canal has the potential to be different both in relative magnitude and time delay for the one-earphone and two-earphone presentations. Accordingly, when the third presentation was subtracted from the sum of the first two presentations, there was the potential for a phase interaction between the two interim resultants that could be either constructive or destructive in its effect on the final resultant. If the interaction was constructive during the inattention condition, it might continue to be constructive during the attention condition or it might change to destructive interference, depending upon the time delays involved. Thus, if the stiffness and compression changes were greater for the attention conditions than for the inattention conditions (as we are assuming), then the difference in nSFOAE magnitude could be either positive or negative depending upon these phase effects for the ear in question. Kujawa and Liberman (2001) also appealed to constructive and destructive interference to explain different directions of effect across ears and across stimulus parameters in another form of OAE; the general topic was discussed by Russell and Lukashkin (2008).

In an attempt to make the preceding argument about phase more visual, we offer Fig. 9, which contains phasor diagrams for inattention and attention conditions. The top panel illustrates hypothetical conditions for an inattention condition, and the bottom panel illustrates hypothetical conditions for attention. Phasor length represents the *amplitude* of a response, but for simplicity, we will talk in decibel units. (Phasors are appropriate here because the preponderance of each response is to the 4.0-kHz tonal stimulus.) In both panels of Fig. 9, the thin solid lines at the bottom represent the sum of the *responses* (i.e., SFOAEs) to the two single-earphone presentations of a triplet. The heavy solid line in the top panel represents the response to the two-earphone presentation of that triplet during inattention. The *stimulus* for the two-earphone presentations was essentially 6 dB greater than each of the stimuli for the single-earphone presentations, and we assume that that 6-dB difference is responsible for both additional amplitude compression and a reduced time delay for the two-earphone response compared to the two single-earphone responses. Thus, this two-earphone response is shown as shorter by 2 dB, and as having a 15° phase shift, compared to the sum of the two single-earphone responses. The dashed line in the top panel represents the nSFOAE—i.e., the "difference waveform" after the two-earphone response has been subtracted from the sum of the two single-earphone responses. The length and relative phase of that phasor will be used as the reference for inattention against which the phasor constructions for attention will be compared.

The bottom panel of Fig. 9 contains three phasor constructions for the attention condition that are meant to illustrate individual differences in the phase shift between the responses for the single- and two-earphone presentations. Also, the lengths of the phasors for the single- and two-earphone presentations are each about 1 dB shorter than for the inattention condition in the top panel. This decrease in response magnitude is attributable to the greater efferent activity we assume to exist in all subjects during attention. As the figure illustrates, when the phase shift between the single- and two-earphone responses is 15°, as in the top
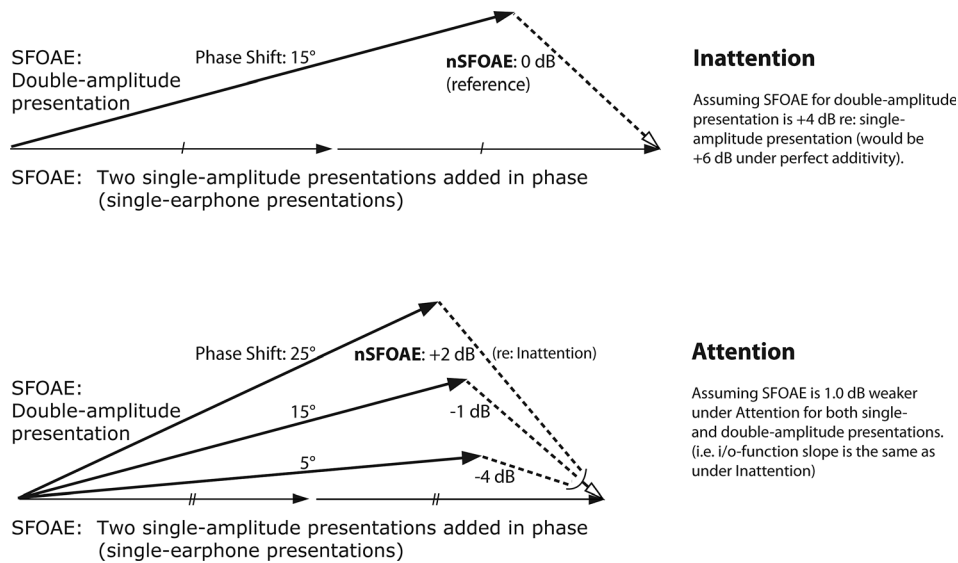
FIG. 9. Phasor diagrams for an Inattention and Attention condition. (Top) Inattention condition: When the response to the two-earphone presentations is subtracted from the sum of the two single-earphone responses, an nSFOAE is left as a resultant (dashed line). Call that the reference (0 dB) for comparison with the attention condition. (Bottom) Attention condition: Assume that efferent activity always is stronger during attention than inattention, so that the strength of all responses is slightly weaker than in the top panel. Also assume that the phase angle (time delay) between the response to the two-earphone presentations and the sum of the responses to the two single-earphone presentations can be different in different subjects. When that phase angle is 5°, the nSFOAE (the dashed lines) during attention is 4 dB weaker compared to inattention; when that phase angle is 15°, the nSFOAE during attention is 1 dB weaker compared to inattention; when that phase angle is 25°, the nSFOAE during attention is 2 dB *stronger* compared to inattention. Thus, even though efferent activity itself always is stronger during attention than inattention, the sign of the difference between attention and inattention can be positive or negative depending upon the time delays in individual cochleas.

panel, the nSFOAE (the dashed line) is about 1 dB weaker than the nSFOAE for the inattention condition. When that phase shift is only 5° (10° less under attention than inattention), that difference in nSFOAE magnitude is even greater (−4 dB). When that phase shift is 25° (10° greater under attention than inattention), however, the magnitude of the nSFOAE for attention is about 2 dB *greater* than that for inattention. (At 4.0 kHz, 5°, 15°, and 25° correspond to about 3, 10, and 17 µs, respectively.) Thus, different subjects could exhibit different directionalities of effect between inattention and attention conditions by having only slightly different phase angles (time differences) between their responses to the single- and two-earphone presentations. We emphasize that this demonstration assumes that efferent activity *always* is greater under attention than inattention, yet the magnitude of the nSFOAE can be larger or smaller for inattention. All of the parameter values selected for this example are regarded to be plausible, but a wide range of other values yields the same conceptual result. Note that for both the inattention and attention conditions, the phasors are "normalized" at 0° in the single-earphone presentations. There may well be additional phase shifts in the single-earphone response for attention vs inattention, but our analyses do not involve any vector (phase-dependent) calculations between attention and inattention responses; comparisons involve only magnitudes calculated *within* each condition.

In summary, we believe that logic, intuition, and parsimony argue that the individual differences in directions of effect for attention seen in Figs. 4 and 6 are attributable to a secondary factor (such as relative phase), not to some subjects actually having greater MOC activity under attention and others having lesser MOC activity under attention.

B. MOC vs non-MOC effects

Because we believe that attention primarily affects the reflexive response of the MOC efferent system, here we have emphasized the segment of the nSFOAE response seen during the tone-plus-noise portion of the eliciting stimulus rather than the segment of the response seen during the tone-alone portion. We believe that the differences between inattention and attention reported here can be attributed to the reflexive MOC system having been differentially activated during inattention and attention by the efferent flow from auditory cortex to superior olive. That is, we believe that cortico-olivary efferent pathways are able to modulate the level of activity in the MOC reflex.

Specifically, we believe that the cortico-olivary efferent flow is greater during attention than during inattention, and that makes MOC strength greater, cochlear-amplifier strength weaker, and amplitude compression less than during inattention. That is, the cochlear operating characteristic is more linear under attention. As discussed above, this sequence of events can produce nSFOAEs that are either stronger or weaker than those during inattention, depending upon the specific values of the time delays involved. But no matter that direction of effect, our data are not inconsistent with that view that, under attention, the MOC efferents are more active and OHC amplification is weaker in all ears.

Note that some subjects did show differences between inattention and attention during the tone-alone segments of the nSFOAE response even though we assume that those tone-alone responses are primarily attributable to nonlinear *mechanical* mechanisms, not neural mechanisms (see above). We suspect that these tone-alone differences are attributable to differential carryover of efferent activity from triplet 1 to

triplet 2, and from trial to trial. The MOC system is sluggish at resetting. Short-term MOC efferent effects can persist for several hundred milliseconds after the cessation of the MOC-eliciting stimuli (Backus and Guinan, 2006; Goodman and Keefe, 2006; Walsh et al., 2010a,b), and longer-term effects also have been reported (Backus and Guinan, 2006; Cooper and Guinan, 2003; Sridhar et al., 1995). The 500-ms intervals separating our nSFOAE-eliciting stimuli ordinarily should have been sufficient for the short-term persistence to dissipate prior to the next eliciting stimulus (Backus and Guinan, 2006). However, logically, the speech or SSN sounds that followed every nSFOAE-eliciting stimulus in some conditions may have prolonged the persistence so that there was carry-over of the MOC efferent effect across the presentations of each triplet and across the two triplets of each trial. Thus, each stimulus presentation of each triplet may not have started from a state of no efferent activity for all subjects in all conditions. No attempt was made to measure such carry-over effects, but some of the individual differences observed in nSFOAE morphology may be attributable to such carry-over effects across presentations. In this context, note that in the visual-attention study, one pair of conditions had no speech or SSNs to prolong the efferent effect yet those attentional differences (for both tone-alone and tone-plus-noise) were much like those present in the auditory and visual conditions having speech or SSNs (see Table II).

Recently, we also appealed to short-term persistence of MOC activity to explain the attentional differences in nSFOAE magnitude seen during the 30-ms silent period that followed the offset of the tone-plus-noise stimulus (Walsh et al., 2014a,b). That should be thought of as an early stage of the (sustained) persistence we are hypothesizing here to have augmented the attentional differences during tone-alone. During the silent period, unlike the perstimulatory period reported here, the directionality of effect was the same for all subjects: The magnitude of the response was unambiguously greater for inattention than attention (Walsh et al., 2014a,b). The fact that all subjects showed the same direction of effect during the silent period, where phase interactions between the tonal stimulus and evoked OAE were not relevant, supports the notion of a common underlying physiological mechanism that operates differentially during periods of attention versus inattention.

C. Auditory and visual attention compared

The results from the auditory- and visual-attention studies were similar in many ways. At 4.0 kHz, the differences between inattention and attention generally were larger for tone-plus-noise than for tone-alone, the tone-plus-noise differences generally were larger for triplet 1 than triplet 2, and the directions of effect differed across subjects and across comparisons. In the visual-attention study, the differences between inattention and attention (for tone-plus-noise) generally were greater when speech or SSNs were presented as additional distracters than when they were not, suggesting that these auditory distracters did make the visual digit-identification task more difficult. Furthermore, the fact that there were moderate differences between inattention and

attention even when no speech or SSNs were presented is revealing about the functioning of both the cortico-olivary and MOC efferent systems during non-auditory attention.

We admit that some of the experimental similarities between our auditory and visual studies were superficial. The competition between two simultaneous talkers, whether dichotic or diotic, surely was much greater than the (limited) competition between two strings of visual digits contained in colored boxes separated by several degrees of visual angle. Even so, similar results were obtained in the visual-attention study whether or not simultaneous speech or SSNs were present. Apparently there is greater cortico-olivary and MOC activity even when attention is being expended on a *visual* task, just as was suggested by Hernández-Peón et al. (1956) [based as it was on experimentally flawed work (Worden, 1973)]. Learning whether and how visual attention is improved by this auditory activity stands as a challenge for future research. Whether differential MOC efferent activity exists when attention is extended to other sensory modalities is yet to be determined.

D. Ear suppression

An interesting outcome to emerge from the dichotic auditory-attention condition was that there was no difference between the nSFOAE responses obtained from the ear to which the subject was attending and the opposite ear. This absence of "ear suppression" was surprising, in part because the wiring of the OC efferent system is binaurally complex and gives the impression that efferent inhibition could be different for the two ears when that can help perception. We presume that this absence of evidence for ear suppression is at least partly attributable to the specific details of our stimuli and procedure, and that (slightly?) different experimental details might lead to bilaterally asymmetric MOC activity during some dichotic attention task. It must be borne in mind, however, that the attentional effects seen in the visual-attention experiment were very similar to those seen in the auditory experiment. This suggests that, for our experimental procedures, the ear attended to, or even the sensory modality attended to, mattered little. What did seem to matter was the *degree* of attention that was required of the subject; in the visual experiment, attentional effects were greater when an additional distractor was present.

E. Across-frequency effects

Another interesting outcome from both studies is that there were differences between inattention and attention at frequencies far removed from the 4.0-kHz tone. These data were obtained by moving the center frequency of the analysis filter over a range of about 1.0–7.0 kHz (see Fig. 7). The strong implication is that our auditory stimuli (the digits plus the nSFOAE-eliciting stimuli) were producing MOC efferent activity across a wide frequency range. It remains to be determined if other stimuli or tasks lead to less-global, more frequency-specific attentional effects.

2754   J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh et al.: Efferents and attention

F. Possible alternative explanations

A long-standing problem with studies of the human MOC efferent system is that the middle-ear reflex (MER) may have contributed to the outcomes observed (e.g., Goodman and Keefe, 2006; Henin *et al.*, 2014). Thus, a possible alternative explanation for the results reported here is that the MER was differentially activated during attention and inattention conditions. Logically, there appear to be two forms of this alternative explanation. One is that our subjects were *intentionally* contracting their middle-ear muscles (MEMs) and doing so differently in the attention and inattention conditions. Evidence exists that some people are able to contract their MEMs voluntarily (Burns *et al.*, 1993; Reger *et al.*, 1963), but such people are quite rare (about 1% of the population); thus, it is statistically unlikely that all of our subjects possessed this unusual skill. Some people can be trained to voluntarily contract their MEMs (Reger *et al.*, 1963), but no such training was offered to our subjects. Also, there are large individual differences in how long people can maintain their voluntary contractions; each of our blocks of trials lasted between 4 and 6 min.

The second form of this alternative explanation is that the MEMs were contracting reflexively ("unconsciously") but differently for the attention and inattention conditions. That is, the top-down influences that we have been assuming were acting on the MOC network were acting (also or instead) on the MER. We have no direct evidence contradicting this form of the alternative explanation because no simultaneous or auxiliary measurements were obtained of the MER for our subjects (beyond screening for "normal" tympanometry). However, there are several facts that are relevant to the acceptance of this alternative explanation. First, unlike the MOC, we are not aware of any neural-reflex circuitry that could mediate differential MER activity under different states of attention. Second, our most important measurements were made at 4.0 kHz, and the MER primarily attenuates lower frequencies (Dallos, 1973; Goodman and Keefe, 2006; Henin *et al.*, 2014; Schairer *et al.*, 2007). Third, there are large individual differences in the SPLs that activate the MER (e.g., Goodman and Keefe, 2006; Henin *et al.*, 2014), and when a 4.0-kHz tone of 80 dB was presented during the initial screening of these subjects, only three had an MER. By comparison, for our two studies, the 4.0-kHz tone was only 60 dB SPL and our noise only about 63 dB overall. In the past, such levels have been viewed as too weak to activate the MER in most subjects (Dallos, 1973; Keefe *et al.*, 2009; Schairer *et al.*, 2007). Fourth, the rise time of the MER is quite long (Church and Cudahy, 1984; Dallos, 1973; Henin *et al.*, 2014) and the tone-alone period showed no evidence of this onset. Fifth, and most important, the acoustic stimuli presented during the attention and inattention conditions should have been similarly effective at reflexively activating the MER because they had the same levels, durations, and spectral and temporal characteristics; this was true for both the speech (SSN) stimuli and the nSFOAE-eliciting stimuli. All that differed for attention and inattention conditions was the salience of the digit stimuli with regard to the behavioral task. So, if the MER was

activated reflexively in some subjects, the level of activation should have been highly similar across conditions and studies.

Another possible alternative explanation for our experimental differences across attentional conditions is that during the inattention conditions most of our subjects simply became more fidgety and thus more noisy and thus had stronger nSFOAE responses than during the attention conditions. This is an implausible explanation for several reasons. First, our behavioral tasks did require that the subjects attend to the trial-timing sequence during inattention conditions because they had to press a response key at the end of each string of digits, whether auditory or visual. Second, those RTs generally were faster for the inattention conditions than for the attention conditions (this was true for 7 of 8 subjects). Third, the number of trials having no acceptable behavioral response, or having been rejected as noisy or artifact contaminated, was essentially the same for the inattention and attention conditions (see Walsh *et al.*, 2014a,b), an unlikely outcome if the subjects were just more fidgety and generally noisier during the inattention conditions. Finally, this explanation requires that some people exhibited stronger nSFOAEs when they are more fidgety and others weaker nSFOAEs, and that this direction of effect was not consistent across auditory and visual attention.

G. Function of the efferent system

Over the years, reports have emerged linking the cochlear efferent system to masking (Dewson, 1968; Jennings *et al.*, 2011; Kawase *et al.*, 1993; Walsh *et al.*, 2010b), understanding speech in noise (Kumar and Vanaja, 2004), cochlear protection (Liberman *et al.*, 2014; Maison and Liberman, 2000), and other phenomena, and we have been waiting for a convincing, definitive demonstration establishing *which* of these is *the* primary purpose of the cochlear efferent system—*the* reason the efferent system evolved. However, what the elegant work of Guinan and his colleagues has revealed is that *the* purpose of the cochlear efferent system is superordinate to all of these specific functions. Namely, *the* purpose of the MOC system is to continually adjust the amplification provided by the OHCs so that the range of basilar-membrane motion better fits the quite-limited dynamic range of the primary afferent fibers (Smith and Brachman, 1980; Yates *et al.*, 1990; see Fig. S3 in the online supplement for this article). The MOC efferent system is the automatic transmission of the auditory periphery. The contributions that this system makes to masking, frequency resolution, protection, etc., can be viewed as beneficial side-effects (fringe benefits) of this primary purpose of the system.

Auditory scientists long have known that the loss of OHCs carries a "double whammy"; there is both a loss of hearing for weak sounds in those frequency regions having damaged OHCs, and a loss of whatever advantages the cochlear efferent system provides for everyday listening. Following Guinan, the modern take on this second whammy is that the loss of any advantages affecting masking, frequency resolution, cochlear protection, etc., that are

J. Acoust. Soc. Am., Vol. 137, No. 5, May 2015

Walsh *et al.*: Efferents and attention   2755

provided by the normal cochlear efferent system is secondary to the primary consequence of cochlear-efferent loss: Loss of the automatic gain control of basilar-membrane motion.

H. Final comment

Although both the auditory and visual studies demonstrate that the MOC efferent system was differentially active during our attention and inattention conditions, it remains to be demonstrated exactly how additional MOC efferent activity might help with the processing of the auditory or visual stimuli during selective attention.

ACKNOWLEDGMENTS

[1]See supplementary material at http://dx.doi.org/10.15781/T2WC7K for the complete RT data sets for the auditory-attention study; averaged ipsilateral and contralateral responses for all eight subjects and for both triplets; and the complete RT data sets for the visual-attention study.

[2]The values plotted in Fig. 2 and used in all other figures and tables in the paper have been "corrected" to take into account the rise time imposed on each 10-ms analysis window by the 10% bandwidth of the analysis filter (see Walsh *et al.*, 2014a,b). This correction was implemented throughout the paper in order to make all values consistent with the across-frequency data shown in Fig. 8.

Backus, B. C., and Guinan, J. J., Jr. (**2006**). "Time-course of the human medial olivocochlear reflex," J. Acoust. Soc. Am. **119**, 2889–2904.

Brown, M. C. (**2011**). "Anatomy of olivocochlear neurons," in *Auditory and Vestibular Efferents*, edited by D. K. Ryugo, R. R. Fay, and A. N. Popper (Springer, New York), pp. 17–37.

Brown, M. C., Nuttall, A. L., and Masta, R. I. (**1983**). "Intracellular recordings from cochlear inner hair cells: Effects of stimulation of the crossed olivocochlear efferents," Science **222**, 69–72.

Burns, E. M., Harrison, W. A., Bulen, J. C., and Keefe, D. H. (**1993**). "Voluntary contraction of middle ear muscles: Effects on input impedance, energy reflectance and spontaneous otoacoustic emissions," Hear. Res. **67**, 117–128.

Church, G. T., and Cudahy, E. A. (**1984**). "The time course of the acoustic reflex," Ear Hear. **5**, 235–242.

Cohen, J. (**1992**). "A power primer," Psychol. Bull. **112**, 155–159.

Cooper, N. P. (**2004**). "Compression in the peripheral auditory system," in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 18–61.

Cooper, N. P., and Guinan, J. J., Jr. (**2003**). "Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity," J. Physiol. **548**, 307–312.

Cooper, N. P., and Guinan, J. J., Jr. (**2006**). "Efferent-mediated control of basilar membrane motion," J. Physiol. **576**, 49–54.

Dallos, P. (**1973**). *The Auditory Periphery: Biophysics and Physiology* (Academic Press, New York), pp. 83–126.

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.

Dewson, J. H. (**1968**). "Efferent olivocochlear bundle: Some relationships to stimulus discrimination in noise," J. Neurophysiol. **31**, 122–130.

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.

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (**2007**). "Auditory attention—Focusing the searchlight on sound," Curr. Opin. Neurobiol. **17**, 437–455.

Froehlich, 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.

Froehlich, 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.

Goodman, S. S., and Keefe, D. H. (**2006**). "Simultaneous measurement of noise-activated middle-ear muscle reflex and stimulus frequency otoacoustic emissions," J. Assoc. Res. Otolaryngol. **7**, 125–139.

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.

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.

Henin, S., Long, G. R., and Thompson, S. (**2014**). "Wideband detection of middle ear muscle activation using swept-tone distortion product otoacoustic emissions," J. Acoust. Soc. Am. **136**, 272–283.

Hernández-Peón, R., Scherrer, H., and Jouvet, M. (**1956**). "Modification of electric activity in cochlear nucleus during 'attention' in unanesthetized cats," Science. **123**, 331–332.

Jennings, S. G., Heinz, M. G., and Strickland, E. A. (**2011**). "Evaluating adaptation and olivocochlear efferent feedback as potential explanations of psychophysical overshoot," J. Res. Assoc. Otolaryngol. **12**, 345–360.

Kawase, T., Delgutte, B., and Liberman, M. C. (**1993**). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," J. Neurophysiol. **70**, 2533–2549.

Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Keefe, D. H., Schairer, K. S., Ellison, J. C., Fitzpatrick, D. F., and Jesteadt, W. (**2009**). "Use of stimulus-frequency otoacoustic emissions to investigate efferent and cochlear contributions to temporal overshoot," J. Acoust. Soc. Am. **125**, 1595–1604.

Kemp, D. T. (**1980**). "Towards a model for the origin of cochlear echoes," Hear. Res. **2**, 533–548.

Kujawa, S. G., and Liberman, M. C. (**2001**). "Effects of olivocochlear feedback on distortion product otoacoustic emissions in guinea pig," J. Assoc. Res. Otolaryng. **2**, 268–278.

Kumar, U. A., and Vanaja, C. S. (**2004**). "Functioning of olivocochlear bundle and speech perception in noise," Ear. Hear. **25**, 142–146.

Liberman, M. C., Liberman, L. D., and Maison, S. F. (**2014**). "Efferent feedback slows cochlear aging," J. Neurosci. **34**, 4599–4607.

Lukas, J. H. (**1980**). "Human auditory attention: The olivocochlear bundle may function as a peripheral filter," Psychophysiol. **17**, 444–452.

Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701–4707.

Maison, S. F., Micheyl, C., and Collet, L. (**2001**). "Influence of focused auditory attention on cochlear activity in humans," Psychophysiology **38**, 35–40.

Meric, C., and Collet, L. (**1992**). "Visual attention and evoked otoacoustic emissions: A slight but real effect," Int. J. Psychophysiol. **12**, 233–235.

Meric, C., and Collet, L. (**1994**). "Differential effects of visual attention on spontaneous and evoked otoacoustic emissions," Int. J. Psychophysiol. **17**, 281–289.

Mulders, W. H. A. M., and Robertson, D. (**2000a**). "Effects on cochlear responses of activation of descending pathways from the inferior colliculus," Hear. Res. **149**, 11–23.

Mulders, W. H. A. M., and Robertson, D. (**2000b**). "Evidence for direct cortical innervation of medial olivocochlear neurons in rats," Hear. Res. **144**, 65–72.

Puel, J.-L., Bonfils, P., and Pujol, R. (**1988**). "Selective attention modifies the active micromechanical properties of the cochlea," Brain Res. **447**, 380–383.

Rasmussen, G. L. (**1946**). "The olivary peduncle and other fiber projections of the superior olivary complex," J. Comp. Neurol. **84**, 141–219.

Rasmussen, G. L. (**1953**). "Further observations of the efferent cochlear bundle," J. Comp. Neurol. **99**, 61–74.

Reger, S. N., Menzel, O. J., Ickes, W. K., and Steiner, S. J. (**1963**). "Changes in air conduction and bone conduction sensitivity associated with voluntary contraction of middle ear musculature," in *Middle Ear Function Seminar*, edited by J. L. Fletcher (U.S. Army Medical Research Laboratory, Fort Knox, KY), Vol. 576, pp. 171–180.

Russell, I. J., and Lukashkin, A. N. (**2008**). "Cellular and molecular mechanisms in the efferent control of cochlear nonlinearities," in *Active Processes and Otoacoustic Emissions*, edited by G. A. Manley, R. R. Fay, and A. N. Popper (Springer, New York), pp. 343–379.

Schairer, K. S., Ellison, J. C., Fitzpatrick, D., and Keefe, D. H. (**2007**). "Wideband ipsilateral measurements of middle-ear muscle reflex thresholds in children and adults," J. Acoust. Soc. Am. **121**, 3607–3616.

Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944–966.

Smith, R. L., and Brachman, M. L. (**1980**). "Operating range and maximum response of single auditory nerve fibers," Brain Res. **184**, 499–505.

Sridhar, T. S., Liberman, M. C., Brown, M. C., and Sewell, W. F. (**1995**). "A novel cholinergic 'slow effect' of efferent stimulation on cochlear potentials in the guinea pig," J. Neurosci. **15**, 3667–3678.

Walsh, K. P., Pasanen, E. G., and McFadden, D. (**2010a**). "Properties of a nonlinear version of the stimulus-frequency otoacoustic emission," J. Acoust. Soc. Am. **127**, 955–969.

Walsh, K. P., Pasanen, E. G., and McFadden, D. (**2010b**). "Overshoot measured physiologically and psychophysically in the same human ears," Hear. Res. **268**, 22–37.

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.

Worden, F. G. (**1973**). "Auditory habituation," in *Habituation*, edited by H. Peeke and M. Herz (Academic Press, New York), pp. 109–137.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–220.