The Dissertation Committee for Kyle Patrick Walsh Certifies that this is the approved version of the following dissertation:

Nonlinear Cochlear Responses Differ
During Selective and Inattentive Listening

Committee:

Dennis McFadden, Supervisor

Craig A. Champlin

Wilson S. Geisler, III

Mary M. Hayhoe

Arthur B. Markman

Nonlinear Cochlear Responses Differ
During Selective and Inattentive Listening

by

Kyle Patrick Walsh, B.S., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin
May 2012

Dedication

This work, and all subsequent work, is dedicated to Dennis McFadden.

I am honored to have committed my life to making you famous.

Acknowledgements

Nonlinear Cochlear Responses Differ
During Selective and Inattentive Listening

Kyle Patrick Walsh, PhD

The University of Texas at Austin, 2012

Supervisor: Dennis McFadden

Previous studies have demonstrated that the magnitudes of otoacoustic emissions (OAEs) measured during behavioral tasks requiring attention were different from OAE magnitudes measured during tasks that did not require attention. The implication of these results 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 always have been small, and comparisons across studies have been made difficult by significant procedural differences. In this study, we used a nonlinear version of the stimulus-frequency OAE (SFOAE), called the nSFOAE, to measure cochlear responses from human subjects while they simultaneously performed behavioral tasks requiring selective auditory attention (dichotic or diotic listening), or relative inattention. The results indicated statistically significant systematic differences in nSFOAE magnitudes within subjects, but opposite directions of effect across subjects. Differences in the nSFOAEs measured during the dichotic- and diotic-listening conditions suggested that cochlear responses differed in these two types of selective-listening tasks. Time constants measured from functions fitted to the nSFOAE responses indicated faster efferent effects during selective listening than during inattentive listening for the majority of subjects. Furthermore, measurements in brief silent intervals

after the nSFOAE stimuli indicated weaker residual nSFOAE magnitudes during selective listening than during inattentive listening for every subject. This collection of findings supports the hypothesis that auditory attentional demands can modulate cochlear processing, presumably to the benefit of the listener.

Table of Contents

List of Tables

List of Figures

Introduction

The medial olivocochlear bundle (MOCB) is the neural efferent auditory pathway that connects the superior olive in the brainstem to the bases of the outer hair cells (OHCs) in the cochlea (Warr and Guinan, 1979). This neural pathway, often described as a negative feedback loop, has been shown by previous investigators to be activated passively in noisy backgrounds, and to improve signal-to-noise ratios for transient signals presented in noise—by decreasing the gain of the cochlear-amplifier system (e.g., Kawase et al., 1993; Kawase and Liberman, 1993). Although the reflexive (passive) properties of the MOCB have been relatively well established, much speculation has surrounded the possibility that this efferent pathway to the cochlea can be modulated by the cognitive and perceptual demands that are imposed upon an individual at any given time.

In a seminal study, Hernández-Peón (1956) reported in cat that the magnitudes of gross electrical potentials recorded from the dorsal cochlear nucleus (DCN) in the auditory brainstem were reduced when the animals interacted with visual, somatic, or olfactory stimuli. DCN-response magnitudes were larger preceding the multi-modal interactions, and increased to their previous values following the interactions. A traditional interpretation of this result is that there is a limited capacity of attentional resources that can be allocated at one time, and therefore unattended perceptual stimuli are inhibited in the peripheral stages of processing. However, a more recent interpretation suggests that selective attention is essential to perform any behavioral task, independent of the available resources (Geisler and Cormack, 2011). Ultimately, this remarkable finding was denigrated on account of poor experimental control, but since this influential and controversial report, there have been numerous attempts to support the finding that the attentional demands of a behavioral task can modulate the responses of the peripheral auditory system, either in the responses of auditory brainstem (as in Hernández-Peón), or in the responses of the cochlea (a review of this literature is provided below). However, after more than one-half century of research, there remains no clear narrative for how cognitive processes—such as the selective allocation of attentional resources—can affect perceptual processing in the auditory periphery.

If attentional demands (or other cognitive or perceptual demands) were capable of modulating afferent auditory responses at the level of the cochlea, the MOCB would be the neural pathway through which attention would exert its effects. This pathway terminates at the OHCs, which are part of the cochlear-amplifier system (Davis, 1992) that is thought to be involved in the production of otoacoustic emissions OAEs—weak sounds produced in the inner ear and measured in the external ear canal (Kemp, 1980). For this reason, OAEs have been used in humans to examine how cochlear responses are affected by MOCB activation, and in turn how attentional demands affect efferent feedback. Previous studies have demonstrated that OAE magnitudes measured during auditory- or visual-attention tasks were different from OAE magnitudes measured during tasks that did not require attention (see below for a detailed discussion of these studies), but the directions of the attentional effects have been inconsistent, the magnitudes of the observed differences always 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 in humans during tasks that required selective auditory attention (dichotic or diotic listening), or relative inattention, and that yielded concurrent behavioral data in the form of response accuracy and reaction time. That is, physiological and behavioral auditory measures were obtained simultaneously. The nSFOAE was used previously to measure cochlear responses to stimuli used commonly in psychophysical tasks (Walsh et al., 2010), and it was used here to measure cochlear responses during listening conditions that were believed to have differed in the amount of attention required to complete the associated behavioral task.

In the next sections, a review of the structure and function of the efferent auditory pathways will be provided, followed by a review of the studies that used OAEs to investigate the effects of attention on the auditory periphery.

EFFERENT AUDITORY PATHWAYS

Warr and Guinan (1979) delineated two anatomically distinct efferent auditory pathways in cat, both projecting from the superior olivary complex in the brainstem to the cochlea. These projections form the olivocochlear bundle (OCB). The lateral superior olivary (LSO) nucleus sent projections that terminated at eighth-nerve fibers near the bases of the inner hair cells (IHCs), primarily in the ipsilateral cochlea. However, it was shown for the first time that some LSO neurons projected to the contralateral cochlea (were "crossed"). Occasional innervations with outer hair cells (OHCs) also were observed, but the authors suggest that these connections may be aberrant. In contrast, projections from the medial superior olivary (MSO) terminated at the bases of OHCs, primarily in the contralateral cochlea. However, there were enough projections from the MSO to IHCs (10% and 12% ipsilaterally, and 3% and 12% contralaterally, in two cats) that the authors could not rule out a legitimate and purposeful neural connectivity. Morphologically, LSO efferent neurons have small cell bodies and unmyelinated axons, whereas MSO efferent neurons have relatively large cell bodies and myelinated axons (Guinan et al., 1983).

The anatomical distinctions between LSO and MSO neurons and neuronal pathways seem to imply a functional distinction. Unfortunately, relatively little is known about LSO efferent neurons except that they produce both excitation and inhibition in the cochlea, and have very slow time constants, on the order of minutes (Guinan, 2006). In contrast, the functional role of MSO efferent neurons is more clear. The connection from the MSO to the OHCs is inhibitory, and activation of the medial olivocochlear bundle (MOCB) decreases OHC activity. The OHCs contribute little to afferent auditory transmission, but they have the unique ability to modify the local mechanical properties of the cochlear partition, thus affecting the form / rate of afferent information sent by IHCs. The OHCs cells are electromotile—they elongate when hyperpolarized and contract when depolarized, essentially exaggerating the displacement of the basilar membrane produced by passive cochlear mechanisms (Brownell, 1985). Because of their role of providing active mechanical feedback, the OHCs can be thought of as cochlear amplifiers (Davis, 1983). The OHCs appear to be responsible for the impressive dynamic range of the human auditory system through a

3

selective amplification of weak incoming sounds and a compression of stronger incoming sounds, and also for the sensitivity to sounds in the quiet, and the sharp tuning of auditory filters. When the MOCB is activated, and OHC activity inhibited, the dynamic range is reduced, IHC thresholds are raised, and tuning becomes broader.

Effects of efferent feedback have been observed with a variety of physiological techniques, both in humans and in non-human mammals. For example, Galambos (1956) demonstrated that shocks to the medulla in cat suppressed auditory-nerve responses to clicks measured in the quiet. Later, Guinan and Gifford (1988a; 1988b; 1988c) performed a series of experiments meant to disentangle efferent effects by selectively stimulating MOC neurons alone (at their origin), or a combination of MOC and LOC neurons (at the decussation of the crossed olivocochlear bundle) while measuring auditory nerve responses in cat. Efferent stimulation shifted rate-level functions to stronger input levels and lowered the rate at which nerve impulses saturated (1988a), depressed spontaneous firing rates (1988b), and both raised the thresholds of tuning curves in the quiet, and caused the tuning curves to be broader (1988c). These effects were similar whether only MOC neurons or a combination of MOC and LOC neurons were stimulated, but were slightly larger using the combinatoric stimulation.

Also in cat, Kawase et al. (1993) and Kawase and Liberman (1993) showed that MOCB stimulation *enhanced* auditory-nerve responses to tones in noise, and compound action potentials (CAP) to tones in noise, respectively, when a broadband noise was presented simultaneously to the contralateral ear. These results appear paradoxical because activation of the MOCB has inhibitory effects, not excitatory. The ipsilateral noise adapted the responses of auditory-nerve fibers by depleting the amount of available neurotransmitter (Smith and Zwislocki, 1975; Smith, 1979), so when the MOCB was activated with contralateral noise, and inhibition was initiated, the magnitude of this adaptation decreased. As a result, when tones were presented in the background noise, more neurotransmitter was available to the neurons most sensitive to the frequencies of those tones. Firing rates increased in response to the masked tones, and decreased in response to the masking noise during MOCB stimulation. The authors called this phenomenon "antimasking." Because we do not commonly listen to

sounds in otherwise silent backgrounds, the authors suggested that a major functional role of the MOCB is to enhance detection of transient sounds in noisy backgrounds, perhaps sounds such as speech formants. The superior olive—the origin of the MOC and LOC efferents—is arranged tonotopically (Robertson et al., 1987), maintaining the possibility that the effects of antimasking could be frequency-specific.

The efferent auditory system described above is only part of the story, however. Just as the superior olivary complex (SOC) sends projections to the cochlear hair cells, the SOC receives efferent projections from the inferior colliculus in the brainstem, as well as direct projections from auditory cortex (Mulders and Robertson, 2000a, 2000b). The very existence of this latter connection invites the question of whether cognitive processes originating in auditory cortex may affect the processing of the sounds upon which those cognitive processes are being based. And because cortex and cochlea are connected indirectly, perhaps the sensory receptors themselves—the cochlear hair cells—are affected by cognitive constructs such as attention. If this were the case, the MOCB appears to be the primary neural pathway through which attention would have the potential to modulate cochlear activity.

Interestingly enough, researchers have commonly used visual attention, or attention to another sensory modality, to study the effects of attention on auditory physiology. The general hypothesis was that neural signals to unattended stimuli, such as those used to evoke the auditory responses, would be attenuated by a neural mechanism that is activated by attention to another stimulus, impinging upon another sensory pathway. One problem with a cross-modal experimental paradigm, however, is that it is nearly impossible to ensure that the subject does not "sneak an occasional 'listen' or auditory peek" during the attentional task (Fritz et al., 2007). For this reason (and perhaps others), other researchers have studied attentional effects on auditory physiology using auditory-attention tasks. Using a within-modality paradigm facilitates the opportunity to examine how attention to one component of an auditory stimulus can affect the neural (and in turn mechanical, as we shall see) processing of that component, and the processing of those components that are ignored.

MEASURES OF AUDITORY ATTENTION

The idea that afferent auditory information may be modified by an efferent attentional mechanism is credited to Hernández-Peón et al. (1956), who measured gross electrical potentials from the dorsal cochlear nucleus (DCN) in cat while the animals interacted with visual, somatic, or olfactory stimuli. The visual stimulus was two mice in a bottle, the somatic stimulus was a shock to the paw, and the olfactory stimulus was the smell of fish. Before, during, and after the cats' interactions with these stimuli, clicks were played from a speaker in the free field to elicit action potentials in the afferent auditory pathway. Compared to DCN responses prior to stimulus presentation, when the cats attended to any of the three non-auditory stimuli, the click-evoked responses were "markedly reduced" and at times "practically abolished". After the competing stimuli were removed, DCN responses increased to their previous magnitudes. The authors conceived of a parsimonious balance between all sensory systems whereby unattended afferent signals are reduced prior to reaching their respective destinations in cortex (the cortex has enough to do).

At the time, these results apparently were controversial (due to a poorly controlled auditory stimulus, no control of motor behavior, etc.), and conflicting reports later followed, suggesting that attention operates in cortex, and not in the periphery (e.g., Picton et al., 1971; Picton and Hillyard, 1974; Collet and Duclaux, 1986; Connolly et al., 1989). Still, the idea was out there, and work continued, and more than five decades later, questions and controversies still remain about how and where [at which synapse(s)] filtering by attention takes place. At present, it has been reported that attention can modulate neural activity as early in the afferent auditory pathway as the cochlear nerve and the cochlear hair cells (reviewed below).

In humans, two predominant physiological measures have been used to study attentional effects on the auditory periphery, the first being auditory evoked potentials (AEPs). AEPs are electrical impulses from the brain in response to sound that are recorded using scalp electrodes (Hall, 1992). As an example, Lukas (1980) measured auditory brainstem responses (ABRs) in sixteen subjects using series of brief 1.0-kHz tone pips as the evoking stimulus. The dependent variables were the amplitudes and latencies of the Wave-V components of the ABR responses, which were compared

across two experimental conditions. In the so-called "listen" condition, subjects were instructed to listen to the evoking stimulus for the duration of the recording epoch. In the so-called "look" condition, subjects performed a visual-detection task during the recording epoch in which they counted the number of times a target letter (Q) appeared on a screen. Five letters (O, P, Q, R, S) were presented randomly, one at a time, at a rate of two letters per second, so the subjects had to concentrate (pay attention) in order to perform the task well. Subjects responded at the end of the condition.

The data showed that the amplitudes of Wave V were reduced during the "look" condition (mean = .64 µV) compared to those measured during the "listen" condition (mean = .76 µV), and this difference was statistically significant. Longer Wave V latencies also were observed during the "look" condition (mean = 7.35 ms) compared to the "listen" condition (mean = 7.32 ms), and this difference also was statistically significant. The authors concluded that the responses of the inferior colliculus, which is thought to be the source of the Wave V component of the ABR, can be inhibited by visual attention in situations where sound is irrelevant or detrimental to performance in a task. In a follow-up experiment that used higher-frequency tone pips (2.0 and 8.0 kHz) to evoke a distinct auditory-nerve response (Wave I of the ABR), Wave I amplitudes were significantly smaller, and latencies were significantly longer in the "look" condition compared to the "listen" condition, and this was true for both frequencies. The implication was that even the first synapse in the afferent auditory pathway is inhibited during tasks that require visual attention. Wave V amplitudes and latencies showed the same direction of effect as a function of condition, but the observed differences were statistically significant only for the 2.0-kHz tone.

The results of this study provided, "the first evidence from human subjects that concentrated visual attention can modify the sensory transmission properties of the human auditory nerve," but this was not the first attempt to find a peripheral attentional filter. Lukas suggests that several aspects of his procedure allowed him to observe attentional effects where others had failed (cf. Picton et al., 1971; Picton and Hillyard, 1974). First, he used short trials, each 60 s in duration, so that subjects were not forced to maintain attention for protracted periods of time. Second, each trial was initiated by the subject so in effect *they* decided when they were ready to concentrate on the task.

Third, the visual task was designed to be challenging; said differently, subjects did more than simply look at something. Finally, the frequencies of the tonal elicitors were chosen so as to be within the mid-frequency spectral range where efferent innervation is most dense (Warr and Guinan, 1979). Lukas contended that because the auditory input from a click stimulus contains primarily high-frequency energy, the effects of attention may be lost on click-evoked responses.

A second physiological measure that has been used to study attention and audition in humans is otoacoustic emissions (OAEs). OAEs are weak sounds produced by the inner ear either in response to an acoustic stimulus, or spontaneously (Kemp, 1980). Because OAEs are generated at least in part by the OHCs, they are thought to provide a unique window unto the active processes of the cochlea (which are known to be affected by efferent activation, but not necessarily by attention). Compared to AEPs, OAEs appear to have at least two significant advantages as potential indices of auditory attentional effects, and these advantages have been acknowledged by many previous researchers (Puel et al., 1988; Froehlich et al., 1990; Giard et al., 1994; Meric and Collet, 1994a; Meric and Collet, 1994b, etc.). First, OAEs can be elicited using lower input levels than those needed to evoke detectable responses from the brainstem. This may be important because efferent activation has been shown to be strongest at moderate sound pressure levels. Second, less averaging of the evoked responses is required for OAEs compared to AEPs, so data collection takes less time. This means that subjects have to maintain focused attention for shorter durations.

PRIOR INVESTIGATIONS OF ATTENTION AND EMISSIONS

In (what is apparently) the first investigation of attention using otoacoustic emissions as the primary measure, Puel et al. (1988) recorded evoked otoacoustic emissions (EOAEs) before, during, and after a visual detection task (the acoustic stimulus was a click, but only the dominant frequency of each CEOAE response was measured; dominant frequencies differed across subjects, but were constant for a given subject). Subjects were asked to detect a target letter (Q) in a rapidly presented visual display that contained the similarly shaped letters Q and O (task adapted from Lukas, 1980). Subjects were instructed to press a button whenever the target letter appeared,

and to count (and later report) the total number of target appearances during the emission-recording epoch. EOAE responses recorded before, during, and after the detection task were compared individually to a control EOAE response that was measured prior to these three experimental conditions. In the control condition, subjects were not instructed to do anything, nor attend to anything.

On average, EOAE amplitude was reduced *during* the visual task, relative to the control EOAE amplitude, and this reduction was statistically significant. Thirteen of the 16 subjects showed this effect. The data were not analyzed specifically with respect to changes *within* an experimental condition, but the authors observed that during the visual task, EOAE amplitudes decreased quickly, and then remained at an essentially asymptotic amplitude. Conversely, average EOAE amplitudes before and after the visual task were not significantly different from the control EOAE amplitude, and were not significantly different from one another. Said differently, after the visual task, EOAEs returned to amplitudes commensurate with those measured before the visual task. Puel et al. concluded that "concentrated visual attention can modify EOAEs," and that the most likely pathway through which attention was exerting its effects on afferent auditory transmission was the MOCB (see Warr and Guinan, 1979).

Using similar experimental procedures, and having the same empirical motivation, Froehlich et al. (1990) also showed an effect of attention on the amplitudes of EOAEs (in fact, the acoustic stimulus was a click, and the responses were filtered broadly from 500 – 6000 Hz). The click-presentation rate, the number of responses in each EOAE average, the visual-attention task, and even the number of subjects (i.e., the statistical power) were identical those in the Puel et al. (1988) study. In the Froehlich et al. study, however, the control and attentional conditions were repeated with broadband noise presented simultaneously to the contralateral ear at 50 dB SPL. Contralateral noise has been shown to reduce the amplitudes of OAEs. An additional difference was that Puel et al. measured only the dominant frequency in each click-evoked response, whereas Froehlich et al. retained the spectral lionshare (500 – 6000-Hz pass-band) of each click-evoked response.

Compared to the responses measured in the quiet, EOAE amplitudes were significantly lower when contralateral noise was presented simultaneously, and this was

true whether or not subjects were engaged in the visual task. In other words, the contralateral noise suppressed the responses, presumably via MOCB crossed efferent connections. Importantly however, EOAE amplitudes in the visual-task and no-task conditions were not significantly different from one another, and this was true whether or not the contralateral noise was present. Response amplitudes were lower in the visual-task condition for 3 of the 16 subjects, compared to 13 of 16 in the Puel et al. study. Reanalysis of each EOAE response according to the most dominant frequency component (as in Puel et al., 1988) did not change the results.

The same set of conditions was re-run later in Froehlich's two subjects who had initially shown the largest decrements in EOAE amplitude in the visual-task conditions; EOAE amplitude continued to be mediated by experimental condition. For one subject, EOAE amplitude decreased by an average of 0.2 dB over nine consecutive measures (visual-task condition versus no-task condition), and for the other subject, the average decrement was 0.4 dB over seven consecutive measures. Froehlich et al. speculated that not observing consistent attentional effects in a larger subset of their subjects was due to individual differences in efferent activation. They finish by noting that their experimental procedure (attend vs. not attend, essentially) does not necessarily employ *selective* attention, but rather may exploit differences in general physiological arousal. "The effect of selective attention can be demonstrated only in experiments where two or more channels of stimuli are present in the environment and when attention is switched from one to the other."

In a follow-up study, Froehlich et al. (1993a) became one of the first to use auditory stimuli to investigate attentional effects on cochlear processing. They continued to use visual stimuli as well, and reported an interesting inter-modal difference of effect. In 13 subjects, TEOAEs (as before, the acoustic stimulus was a click, and the responses were filtered broadly from 500 – 6000 Hz) were measured in a reaction-time task. In the auditory-attention condition, a 2800-Hz tone was presented to the contralateral ear at 85 dB SPL at random intervals between 0.3 and 4.0 s, and the task of the subject was to press a button as soon as the tone was detected. The task was identical in the visual-attention condition, but the stimuli were flashes of light. In a

control condition, subjects were only presented with the emission-evoking stimulus, and did not perform a motor task. Contralateral noise was not used in this study.

For both the auditory- and visual-attention conditions, TEOAE amplitudes were significantly lower on average than for the control condition. The difference was 0.20 dB for the auditory condition, and 0.15 dB for the visual condition. When the broadband responses were divided into spectral sub-bands, the experimental effect appeared to be modality dependent. In the 960 – 1920-Hz band, TEOAE amplitudes decreased significantly in the visual-attention condition, but not in the auditory-attention condition. Conversely, in the 1920 – 2880-Hz band, TEOAE amplitudes decreased significantly only in the auditory-attention condition. The authors concluded that "visual and auditory attention act on different areas of the cochlear partition." For comparison, Meric and Collet (1992) also found that both visual and auditory attention diminished EOAE amplitudes, although the observed diminution was only statistically significant during their visual-attention task.

Unfortunately, as in their previous study, Froehlich and colleagues were unable to assess whether the significant effects of task on TEOAE amplitudes were due to changes in attention, or were due simply to changes in arousal (changes in arousal *were* presumed to have occurred). For the auditory and visual reaction-time tasks, it could be argued that subjects needed not to employ attentional mechanisms (such mechanisms certainly were not taxed), only passive detection mechanisms. Furthermore, the experimental conditions required a motoric response, whereas the control condition did not. It seems plausible that subject movement and physiological noise would have increased during the button-press responses, potentially interacting with cochlear responses, but no significant difference was found in the intensity of the background noise between the auditory-attention and control conditions (measured in the 1920 – 2880-Hz band); a similar comparison between the visual-attention and control conditions was not reported.

Interestingly, in the same year, Froehlich et al. (1993b) reported that TEOAE amplitudes increased during sleep, and that the ability of contralateral noise to suppress evoked responses was reduced during sleep. This suggests that during periods of rest (and relative quiet) the cochlear-amplifier system (Davis, 1983) is more active than

11

during periods of alertness. That TEOAE amplitudes decreased during an auditory or visual reaction-time task is not incompatible with this finding.

Meric and Collet (1994b) used an interesting combination of measures of spontaneous otoacoustic emissions (SOAEs) and transient-evoked emissions (TEOAEs) to assess attentional effects. Persons with strong and copious SOAEs are believed to have stronger cochlear-amplifier systems than persons with few or no SOAEs, so individual differences in SOAEs may correlate with differential efferent effects on the cochlear-amplifier system. In a first experiment, nine subjects with SOAEs participated in a visual-attention, reaction-time task. Flashes of red light appeared at random intervals between 0.3 and 4.0 s, and the task of the subject was to press a button as soon as the light was detected. SOAEs were recorded before, during, and after the task. No statistically significant experimental effect was found on the frequency (n)or the amplitude of any subject's spontaneous emissions.

In a second experiment, TEOAEs were measured in 13 subjects with SOAEs, and in 12 subjects with no SOAEs, during the same reaction-time task as above. The acoustic stimulus was a click, and the responses were measured in an 80-ms epoch from 2.5 – 82.5 ms post click. The first 20 ms of each recording epoch was used to measure the TEOAE, and the latter 60 ms was used to measure SOAEs, if present. Subjects without SOAEs showed a significant decrease in TEOAE amplitude during the reaction-time task as compared to amplitudes measured before and after the task, which were comparable. The magnitude of the decrease was approximately 0.3 dB. Measures of residual noise in the TEOAE recordings were commensurate across the three conditions. Interestingly enough, subjects with SOAEs did not show the same attentional effect; TEOAE amplitude did decrease during the task on average, but not significantly. Finally, in accord with the first experiment, neither SOAE frequency nor amplitude differed across conditions.

Subjects with SOAEs had average TEOAE amplitudes that were three times larger than the amplitudes from subjects with no SOAEs. This may explain the absence of an observed attentional effect; SOAE energy would have been present in the 20-ms epoch demarcated / set aside for the TEOAE response, thus overwhelming the TEOAE responses, and perhaps any attentional effects imposed on the TEOAE responses.

When SOAE energy is absent, there is seemingly a better opportunity to observe what appear to be very small effects of visual attention. In an interesting comparison, the authors note that the ages of their subjects were significantly negatively correlated with TEOAE amplitudes in those subjects without SOAEs, but were not correlated in those subjects with SOAEs. Here again, SOAE responses may have been dominating the weaker TEOAE responses, thus masking the attentional effect.

At about the same time that the Meric and Collet (1994b) study was being completed, a review article on attention and otoacoustic emissions was being written by those same researchers (Meric and Collet, 1994a). They observed that across studies, there has been general agreement that TEOAEs elicited by clicks are reduced in magnitude during periods of attention to visual or auditory stimuli (cf. Meric and Collet, 1993 for an exception), but that other types of OAEs appear not to be affected by attention. For example, a series of experiments by Avan and Bonfils (1992) showed no effect of attention on active cochlear processing as assessed by distortion-product OAEs (DPOAEs) and stimulus-frequency OAEs (SFOAEs). Furthermore, and as described above, the magnitudes and frequencies of subjects' SOAEs were shown to be essentially constant during a visual-attention task, and TEOAE magnitudes were unaffected by visual attention in subjects with SOAEs (Meric and Collet, 1994b).

Irrespective of whether or not attentional effects on the auditory periphery were observed, the primary criticism of *all* of the studies completed to date was that their experimental procedures did not permit delineation between specific effects of attention, and the more general effects of arousal. In these studies, the common comparison was between OAE amplitude measured during an "attentional" task, and OAE amplitude measured before or after the task was completed. Arousal can be presumed to have increased during the "attentional" tasks, due simply to the elicitation of behavioral responses. Meric and Collet suggested that specific effects of attention can be demonstrated only by using a selective-attention task, in which subjects attend to one stimulus (for the purpose of task performance) while concurrently ignoring other stimuli (that are irrelevant to the task at hand). This type of selective-attention procedure was employed by Giard et al. (1994), and this study is reviewed next.

Giard et al. (1994) showed that the amplitudes of evoked otoacoustic emissions (EOAEs) could be modulated by attention in a dichotic-listening task. As noted above, this study is distinctive compared to those reviewed prior because it was the first experiment of this sort to use a behavioral task in which the listener had to attend to one auditory stimulus, and ignore another that was presented simultaneously (i.e., employ selective attention). The evoking stimuli were pure tones, either 1.0 or 2.0 kHz, and 5 ms in duration. The tones were delivered with equal probability to the two ears, but the same tone always was delivered to the same ear (this was counterbalanced across sessions). The task of the listener was to count the number of "target tones"—tones of slightly greater amplitude—that were embedded in one of the two tone sequences. The listener was instructed to attend either to the left or the right ear, and to ignore the other ear. Thus, the *direction* of attention was the only difference between the attention and the non-attention conditions.

The dependent measure in this study was the difference in EOAE amplitude when the eliciting tones were attended versus when they were unattended, and this comparison was made across sessions, in the same ear. The EOAE responses were analyzed in three adjacent, non-overlapping spectral bands: (1) the "primary band," which was $\pm$ 15% of the tone frequency; (2) the "secondary band", which was adjacent to, and had the same spectral width as the primary band; and (3) the "residual band," which included all other spectral components of the response. Response analysis began 6 ms after the evoking stimulus had decayed in the ear canal, and ended 20 ms after stimulus onset.

Statistical comparisons of mean EOAE amplitudes showed that, in the primary band, there was a small, but statistically significant difference in EOAE amplitude to attended tones versus unattended tones; the evoked responses were larger in amplitude when the eliciting tones were attended. The largest average difference was 0.62 $\pm$ .18 dB. A smaller difference in the same direction was observed in the secondary band of the response, and this difference was statistically significant only in the right ear. Finally, there was no statistical difference in EOAE amplitude in the residual band of the response. The implication of these results is that auditory attentional effects are (at least in some instances) frequency specific. Said differently,

14

only those frequencies close to the attended frequency were meaningfully affected by attention.

As the authors were aware, however, an inherent problem with their experimental procedure was that it did not differentiate whether EOAE amplitude was enhanced when the eliciting tones were attended, or whether EOAE amplitude was attenuated when the eliciting stimulus was ignored and attention was directed to the opposite ear. The most appropriate comparison for these two experimental conditions is a control condition in which the listener is not instructed to attend to any sound, only to reduce their own. Of course, the eliciting stimulus would be presented in this condition, but in the absence of a behavioral task. More recent investigations (see below) have used comparison conditions of this type. Nevertheless, the Giard et al. study did suggest that cochlear processing is modified during tasks that employ selective attention to concurrent auditory stimuli. Curiously, however, the authors did not explain whether their result— an enhancement of EOAE amplitude when the eliciting tones were attended—is compatible (or not) with a host of previous studies showing an inhibition of cochlear responses during attention. To this end, a comparison of the relevant findings from the studies reviewed here is presented below.

There is an important distinction to be made between the studies described previously and those to be described next. Namely, early studies of attention and OAEs commonly measured EOAEs to transient auditory stimuli (clicks and tones) presented in the quiet, but all of the subsequent studies reviewed here measured OAEs using clicks or tones presented to one ear, while broadband noise was presented to the contralateral ear [only one earlier investigation, by Froehlich et al. (1990) also used this arrangement]. Contralateral broadband noise was used in these studies in order to activate the MOCB efferent system; activation of the MOCB inhibits the active processes of the cochlea that are responsible for the production of OAEs. The effects of noise are to reduce the gain of the cochlear-amplifier system, and to make the cochlear input / output function more linear. The principal experimental comparison in these studies was between the amount of suppression imposed on the evoked response by broadband noise when the listener was engaged in an attentional task (auditory, or sometimes visual), and the amount of suppression when the listener was

not instructed to attend to anything. Because the MOCB cannot be ablated in humans for the purpose of empirical investigation, presenting contralateral noise is a convenient and non-invasive way to modulate efferent activity (or at least to ensure that the MOCB is activated to some extent).

In these subsequent studies, another distinction to be mindful of (to attend to) is whether the listeners were attending (or not) to the evoking stimulus, to the contralateral, suppressing stimulus, or to some other sound presented simultaneously with the either the evoking or suppressing stimulus. This distinction can be referred to as the "direction of attention," and will be elaborated upon later when discussing the present research proposal. As an example, the direction of attention in Giard et al. (1994) was toward tones presented to the ear either ipsilateral or contralateral to OAE measurement (the ipsilateral tones were the eliciting stimulus, of course). In the case of a visual-attention task (e.g., Puel et al., 1988), the direction of attention would be toward the visual stimulus upon which behavioral responses were based.

Maison et al. (2001) demonstrated that contralateral suppression of OHC responses by noise can be augmented by auditory attention in a frequency-specific manner. This specificity is an important finding if true because it informs us about the innervation patterns of the human efferent auditory system. Evoked otoacoustic emissions (EOAEs) were recorded in the right, ipsilateral ear of each listener using 1.0- and 2.0-kHz tones that were 4 ms in duration and approximately 60 dB peSPL. For some of the conditions, the tones were presented in the quiet, but for other conditions, broadband noise was presented simultaneously at 30 dB SL to the contralateral ear for the duration of the EOAE recording epoch (to elicit the MOCB and suppress the ipsilateral responses). For some of the noise conditions, 150-ms tone bursts called "cues" were presented simultaneously with the noise in the contralateral ear; these tones were 10 dB weaker than the noise, were 1.0, 2.0, or 4.0 kHz, and occurred randomly within each 1-s recording epoch. For 10% of the epochs, the "cue" tone was followed by a 100-ms "probe" tone of the same frequency.

In the first part of the experiment, EOAEs were recorded in the presence or the absence of the contralateral stimuli. Subjects were asked to remain quiet during the recordings, but were not informed about the nature of the stimuli. These were referred

to as "no task" conditions. In the second part of the experiment, EOAEs again were recorded in the presence or the absence of the contralateral stimuli, but this time the subjects were instructed to count the number of probe tones that were presented with the contralateral noise. These were referred to as "attentional" conditions. The no task conditions always preceded the attentional conditions.

The results showed that all of the contralateral stimuli produced a significant suppression of ipsilateral EOAE amplitude. Furthermore, there was no difference in the amount of suppression caused by the different contralateral stimuli, indicating that the broadband noise acted primarily as the suppressing stimulus, and that the various cue and probe tones had no additional or differential effects. EOAE suppression was significantly greater in the attentional conditions than in the no-task conditions, but *only* when the frequency of the contralateral, attended tones matched the frequency of the ipsilateral, evoking tones. The differences in EOAE amplitude were about 0.2 – 0.3 dB. In other words, these results suggested that attending to a frequency region in one ear can inhibit OHC activity in the comparable frequency region in the opposite ear. Physiologically, this would mean that MOCB fibers innervating attended frequency regions are more active (produce more inhibition of OHC activity) than MOCB fibers innervating unattended frequency regions.

de Boer and Thornton (2007) measured click-evoked otoacoustic emissions (CEOAEs) in human ears in the presence and absence of a broadband noise presented simultaneously to the contralateral ear. The CEOAE responses were measured at two click levels (50 and 60 dB SPL), during three different task conditions: (1) a "passive visual task" condition in which the listener watched a silent movie; (2) an "active visual task" condition in which the listener judged the correctness of simple additive sums; and (3) an "active auditory task" in which the listener detected tone pips (2.0 kHz, 20 ms) embedded in the click-train stimulus. These responses then were compared to CEOAE responses measured during a "no task" condition, in which the listener simply was required to stay awake.

Effect sizes indicating the amount of suppression caused by the contralateral noise were computed for each task. The amount of suppression (about 2 – 3 dB across conditions) was statistically significant for each task, and for the two click levels. For the

attentional effects, CEOAE amplitude was assessed using a repeated-measures analysis of variance (ANOVA), with task (see above), click level (50 and 60 dB SPL), and contralateral noise (presence or absence) as the within-subject factors. There was no main effect of task on CEOAE amplitude when the contralateral noise was present, and this was true at both click levels. In other words, what the listener was doing during CEOAE measurement had no reliable effect on the amount of CEOAE suppression produced by the contralateral noise. Additionally, there was no main effect of task when the contralateral noise was absent, and this too was true at both click levels.

Although task did not affect CEOAE suppression magnitude, task did have a statistically significant effect on "I/O suppression," defined as the change in the slope of the cochlear input / output function caused by the contralateral noise. As mentioned above, broadband noise has been shown to both decrease the gain of the cochlear-amplifier system, and to make its input / output function more linear (less compressive). Only the active auditory task showed no difference in slope when the contralateral noise was present versus absent. This finding implies that when the listener is actively engaged in an auditory task, contralateral noise has less influence on the slope of the cochlear input / output function (however, employing auditory attention did not change suppression magnitude). In other words, active auditory attention was shown to inhibit MOCB suppression. For the other three tasks, the changes in slope calculated between the 50- and 60-dB click levels were about 0.05 – 0.10 dB / dB.

Although the difference was not statistically significant, the active auditory task did show the smallest magnitude of CEOAE suppression of all four tasks, again suggesting that MOCB activity was inhibited when auditory attention was employed. However, the differences in suppression magnitudes from other tasks were just tenths of a decibel. There is no mention of whether additional spectral analyses were performed on the click-evoked responses to examine whether certain components of the responses were suppressed more than others. For example, it would have been interesting to filter the CEOAEs around the frequency of the tone that subjects attended in the active auditory task (using a pass-band similar to Giard et al.'s "primary band"), or to make a gross distinction between the responses at nominally high versus low frequencies. Finally, the de Boer and Thornton result that neither "active" nor "passive"

visual attention changes the cochlear processing of transient sounds contradicts the findings of Puel et al. (1988), Froehlich et al. (1990), Froehlich et al. (1993), and Meric and Collet (1992; 1994b), who demonstrated that emission amplitudes were reduced during tasks requiring visual attention.

In addition to the magnitudes of CEOAE suppression, recording-noise levels also were measured de Boer and Thornton. Replications of each experimental condition allowed derivation of both average and difference waveforms, and the difference waveforms served as estimates of residual, noncochlear noise (the repeatable, presumably cochlear components were canceled in the subtraction). Rejection rates were computed for each condition as well. Compared to the no-task condition, recording-noise levels were significantly weaker in the passive visual task condition, and nearly significant ($p = 0.07$) in the active auditory task condition, suggesting that during certain tasks requiring attention, a listener may be able to reduce the physiological sounds that could interfere with performance. Rejection rates also were significantly lower in the active-auditory and passive-visual tasks compared to the no-task condition. For neither measure, however, was the active visual task condition different from the no-task condition; the former condition required a button press, which was perhaps a source of additional noise. de Boer and Thornton speculated that focused concentration on either auditory or visual stimuli will encourage a subject to suppress controllable noise. If this is the case, differences in recording-noise levels might create small differences in CEOAE response amplitudes, and these small differences might be interpreted as differences in the physiology underlying the phenomena of interest.

Perhaps the most interesting finding the de Boer and Thornton study was an apparent inhibition of MOCB suppression during an "active auditory task." This was the only task for which the slope of the cochlear input / output function did not change significantly in the presence of contralateral noise; for all others tasks the contralateral noise caused a small increase in the estimated slope value. This finding, that auditory attention can inhibit MOCB effects (I/O slope but not suppression magnitude) appears to be at odds with the findings of Maison et al. (2001) in which auditory attention *enhanced* contralateral suppression effects. In that study, subjects detected tone pips embedded in the contralateral ear rather than in the eliciting click train. The two results

are compatible only if auditory attention decreases contralateral suppression in the attended ear (de Boer and Thornton, 2007) and increases suppression in the unattended ear (Maison et al., 2001). Whether the effects of attention can be lateralized in this way remains to be tested specifically.

In a recent study, Harkrider and Bowers (2009) demonstrated that attention to *either* an ipsilateral, emission-evoking click or to a contralateral, broadband noise can decrease the amount of suppression of the click-evoked responses caused by the noise. Click-evoked responses were obtained via a nonlinear summation procedure in which every fourth click was phase inverted, and amplified to a level three times that of the preceding three clicks; sets of four successive clicks (and CEOAEs) were summed, canceling the evoking stimuli and preserving the cochlear responses. This is an interesting procedural difference from the previously discussed studies because it allows examination of the earliest parts of the transient-evoked responses that are otherwise obscured by the (much stronger) evoking stimulus. The authors noted, however, that contralateral broadband noise produces more suppression of low-frequency, longer-latency components of an ipsilateral click-evoked response (e.g., Maison et al., 1997). CEOAE-response storage alternated between two buffers.

The amount of suppression of ipsilateral click-evoked responses by contralateral broadband noise was measured during one non-attention condition, and two attention conditions. The non-attention condition always was completed before the attention conditions, and the order of the attention conditions was counterbalanced; every condition was repeated once. In one attention condition, subjects were instructed to count the number of "ringing pulses" they heard during the ipsilateral click stimuli; a ringing pulse indicated that a CEOAE response was rejected for exceeding a recording-noise criterion. In the other attention condition, subjects were instructed to listen for a low-level speech waveform embedded in the contralateral noise; in fact there was no speech present in the noise. To be clear, the emission-evoking and emission-suppressing stimuli were exactly the same across conditions, and only the presence or absence of attention, and the direction of employed attention were varied. Clicks were presented to the right ear at 12 dB above the level necessary for a just-detectable CEOAE response in the presence of the contralateral noise; across subjects, click

20

amplitudes were between 60 – 77 dB peSPL. The contralateral noise was presented to the left ear at 40 dB SL (55 – 60 dB SPL).

The mean CEOAE amplitude for each of the experimental conditions was subtracted from the mean CEOAE amplitude measured without contralateral noise. CEOAE-response amplitudes in the quiet were 9.2 dB on average, and the average amount of contralateral suppression in the non-attention condition was 1.4 dB (range = 0.2 – 3.1 dB). In order to evaluate the differences in the amount of contralateral suppression across conditions, a one-factor (experimental condition) repeated-measures ANOVA was used. Compared to the non-attention condition, significantly less contralateral suppression was observed in both attention conditions (about 0.5 dB less, on average), suggesting an attention-induced, cortically mediated release from the cochlear inhibition imposed by the MOCB. The amounts of suppression in the two attention conditions were not significantly different, however, suggesting that the auditory attentional mechanisms mediating this release from MOCB inhibition were indifferent to the directionality of the listeners' attention. The active process of the cochlea appears to be most influenced by MOCB inhibition when the listener is not attending to sound, and somewhat less influenced (a release from inhibition) when the listener is attending to sound.

Each CEOAE was analyzed further by parsing the response waveform into adjacent, non-overlapping time intervals (0 – 3 ms, 3 – 6 ms ... 15 – 18 ms). The purpose of this manipulation was to examine whether different spectral components of the CEOAE response were differentially suppressed by the contralateral noise. In both attention conditions and the non-attention condition, the magnitude of CEOAE suppression was significantly less in earlier versus later time intervals, indicating that the low-frequency components of the CEOAE responses were suppressed more than the high-frequency components. The interaction between attentional task and time interval was not significant, indicating that attending (or not) to sound did not change the spectral balance of the CEOAE responses.

Finally, as in de Boer and Thornton (2007), recording-noise levels also were measured as potential indicators of auditory attentional effects. First, the number of rejected CEOAE responses was counted for each subject and condition. Second, an

estimate of non-cochlear, uncorrelated noise in each CEOAE average was measured as the residual difference when the mean CEOAE response from one data buffer (A) was subtracted from the mean CEOAE response from a second buffer (B). During CEOAE measurement, response storage alternated between these two buffers. In contrast to the de Boer and Thornton finding, Harkrider and Bowers found that their recording-noise-level measures were not influenced significantly by experimental condition. The number of rejected CEOAE responses was lower in the attention conditions compared to the non-attention condition, but for all conditions the standard deviation of number of rejections was larger than the mean number of rejections.

For the purpose of providing across-study comparisons, the most important experimental details from the studies reviewed above are presented in Table 1. For each study, information is provided regarding (1) the type of OAE measured, (2) the presence or absence of contralateral noise, (3) the behavioral task of the subject, (4) the focus of attention, and (5) the main findings. A brief interpretation of those findings in terms of the MOCB efferent system, and with respect to other relevant findings, is presented in the subsequent text. This table is intended to provide a concise summary of the most relevant comparisons; additional experimental details can be found in the preceding text.

Examination of Table 1 reveals that across-study comparisons become complicated quickly, because the experimental procedures differed in so many ways. For example, Giard et al.'s (1994) finding that attention to an eliciting tone can *enhance* the cochlear responses to that same tone is intriguing, but it cannot be compared directly to any other findings because those experiments that also measured emissions in the quiet used a visual-attention task, not an auditory task, and they used click stimuli rather than tonal stimuli. The sole exception, Froehlich et al. (1993a), also used an auditory task, but attention was directed only to the contralateral ear, and the result was a decrease in CEOAE amplitude, not an increase. Maison et al. (2001) used tonal elicitors, like Giard et al., but the Maison study is different in that the cochlear responses were suppressed with contralateral noise, and attention was directed away from the eliciting stimulus, towards the opposite ear. Finally, two other studies had subjects attend to an emission-evoking stimulus, as in the Giard et al. study, but one study

22

showed no effect of attention of CEOAE amplitude (de Boer and Thornton, 2007), and the other study showed a significant release from CEOAE suppression (Harkrider and Bowers, 2009); both studies used contralateral noise, however, and click stimuli. Aside from Giard et al., no one has reported an *enhancement* of OAE amplitude during behavioral tasks requiring attention. However, if the differences in EOAE amplitudes observed in the Giard et al. study were due to an attenuation of the responses to unattended tones, rather than an enhancement of the responses to attended tones, then the results are generally compatible with other previous reports showing that cochlear responses to unattended sounds are mitigated by attention to other auditory or visual stimuli (Puel et al., 1988; Froehlich et al. 1990, 1993a; Meric and Collet, 1994b). This interpretation of the Giard et al. result requires that the MOCB can be elicited by attention in the quiet, in the absence of an eliciting sound.

As another onerous comparison, Harkrider and Bowers (2009) found that auditory attention either to an ipsilateral click train or a contralateral broadband noise significantly decreased the amount of CEOAE suppression caused by the noise. The implication of this finding is that attention to sound, irrespective of ear and attentional task, decreases inhibition of the cochlear-amplifier system typically caused by the MOCB in a noisy background (cochlear gain remains high). This result is convincing enough, but the implication requires thought. What would be the benefit of suppressing a mechanism whose activation facilitates the detection of brief sounds in noise, especially when the attentional tasks in this study were to attend either to speech in noise, or to pulses in the click train? In any case, procedural differences again constrain the value of across-study comparisons. Two other studies had subjects attend to the ear in which OAEs were recorded (cf. Giard et al., 1994; de Boer and Thornton, 2007), but all three studies obtained different results (under different conditions). Other studies directed attention to the contralateral, non-recording ear, but only Maison et al. also presented broadband noise to that ear, and they reported an increase in MOCB suppression, not a decrease. In the quiet, Froehlich et al. (1993a) showed that CEOAE amplitude was decreased during auditory attention to the contralateral ear, thus implying activation of the MOCB, not an inhibition of its effects. In fact, only one other study reported what could be interpreted as an inhibition of

MOCB effects during attention. de Boer and Thornton (2007) found that the slope of the cochlear input / output function was unchanged by contralateral noise during auditory attention (MOCB activation normally linearizes the input / output function). Recall, however, that in this study, CEOAE amplitude was unaffected by attention, whether auditory or visual.

As these examples illustrate, it is difficult to draw many substantive conclusions from the research that has been done to date on attentional effects on OAEs. Still, there appear to be important clues in these studies, and the patterns in their data, that ultimately may shed light on the (still uncertain) possibility of attentional regulation of active cochlear processing. First, all four of the studies that employed a visual-attention task reported reduced cochlear responses to unattended, emission-evoking sounds, thus implying that the MOCB was activated by attention in order to mitigate the responses to task-irrelevant auditory stimuli. In three of these studies, emissions were recorded without a broadband noise presented simultaneously to the contralateral ear, suggesting that the MOCB can be activated by attention alone, even if the MOCB has not been activated concurrently by sound. In accord with this interpretation, Giard et al. (1994) showed that in otherwise quiet, EOAE responses to unattended tones were reduced in amplitude compared to the responses to attended tones. Moreover, if it is assumed that attention can *augment* the MOCB effects that can be initiated by sound, then the results of Maison et al. are in agreement with these studies as well. Said differently, the results from the first six studies in Table 1 suggest that visual or auditory attention can activate or augment the effects of the efferent auditory system at the level of the cochlea, if the following assumptions are met: (1) the MOCB can be elicited by attention alone, and (2) attention can act to enhance the MOCB effects that are initiated by sound.

A second potentially important pattern from the reviewed studies is that the sizes of the effects of attention on OAEs are quite small. Puel et al. (1988) found that CEOAE amplitude was reduced by about 1.0 – 1.5 dB during visual attention, when averaged across subjects. In a similar study, Froehlich et al. (1990) found that visual attention reduced CEOAE amplitude by about 0.2 – 0.4 dB in the two subjects who showed the largest effect. In Froehlich et al. (1993a), response diminution was 0.15 dB during

visual attention, and 0.20 dB during auditory attention, on average. In Meric and Collet (1994b) CEOAE amplitude was reduced during visual attention by about 0.3 dB, in subjects without SOAEs. Giard et al. (1994) reported about a 0.6-dB difference in EOAE amplitudes between attended and unattended tone pips. When contralateral noise was presented simultaneously in order to suppress ipsilateral responses, Maison et al. (2001) showed a frequency-specific enhancement of EOAE suppression during auditory attention of about 0.2 – 0.3 dB, while Harkrider and Bowers (2009) showed about a 0.5-dB release from contralateral suppression, which itself was about 1.4 dB on average in the absence of attention. Finally, de Boer and Thornton (2007) reported that auditory attention did not affect CEOAE amplitudes, but rather prevented the noise-activated MOCB from reducing the compressive nonlinearity of the cochlear input / output function; during visual attention, slopes increased by 0.1 dB / dB, on average (input / output functions became more linear).

PRESENT EXPERIMENT

The present experiment evaluated human cochlear responses with respect to concurrent behavioral responses. Viewed broadly, the objective was to explore connections between behavioral and physiological measures of auditory performance during some conditions that required the use of attentional resources, and others that did not. In accord with the lessons of prior research, the present experiment: (1) used behavioral listening tasks that required *selective*-attentional mechanisms, as opposed to general-arousal mechanisms alone (see Giard et al., 1994, for the motivation). In this experiment, subjects were required to do more than simply listen to sounds presented to the ears, or not; rather they were required to make decisions based on the content of the acoustic stimuli; (2) compared residual-noise levels across OAE recordings to ensure that experimental effects were not contaminated by any such differences (e.g., de Boer and Thornton, 2007; Harkrider and Bowers, 2009); (3) used appropriate control conditions with which to compare the attentional conditions, namely conditions in which the listener was not instructed to attend, in which they had not attended to like stimuli previously (e.g., Maison et al., 2001), and in which the motor response was the same as that used for the attentional conditions.

An important distinction from previous studies, however, was that in the present study, subjects were given somewhat challenging behavioral tasks during which their decisions were analyzed for correctness and consistency. In order to perform well, subjects had to attend to certain aspects of the acoustic stimuli while concurrently ignoring other aspects that were irrelevant for the behavioral task (i.e., employ selective attention). (In contrast, in previously-reported attentional tasks, subjects were not assessed in terms of task performance, only task completion.) In addition to having used more difficult listening situations for the attentional tasks in this study, the acoustic waveforms also were different from those used in the studies reviewed above. All previous investigations have used transient stimuli—either brief tones or clicks—as the OAE elicitor, but these stimuli appear not to have been the best choice for studying the effects of the MOCB 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 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 using attention, and whether or not a deliberate MOCB-eliciting stimulus was presented simultaneously with the evoking (and perhaps also MOCB-eliciting) stimulus. This is a problematic possibility. In contrast, Guinan et al. reported that long-duration tones of moderate intensity, like those used to evoke stimulus-frequency OAEs (SFOAEs), were not effective elicitors of the efferent system.

Accordingly, we used a nonlinear version of the SFOAE, called the nSFOAE, that was elicited by a long-duration tone simultaneously presented with wideband noise. The nSFOAE stimulus was not the focus of attention in the conditions used here, but rather was interleaved with the stimulus (speech) waveforms used as the basis for the behavioral tasks. In other words, we used relatively complex speech waveforms to elicit behavioral responses from the listener, and other, relatively simple waveforms to elicit the physiological responses.

Methods

METHODS: GENERAL

In the experiments described here, a nonlinear version of the stimulus-frequency otoacoustic emission (SFOAE), called the nSFOAE, was measured at the same time that subjects performed behavioral tasks requiring selective auditory attention, or no attention. The Institutional Review Board at The University of Texas at Austin approved the procedures described here, and all subjects provided their informed consent prior to any testing. The subjects were paid for their participation.

Two males (both aged 22) and six females (aged 20, 21, 23, 23, 23, and 25) participated in these experiments. 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 four tones to the test ear, each at as many as three levels, while the compliance of the middle-ear system was measured. For the 500- and 4000-Hz tones, the test levels were 80, 90, and 100 dB HL, and for the 1000- and 2000-Hz 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 4000 Hz, the frequency used most commonly in these experiments, three subjects had an MER in either the left or the right ear at 80 dB HL, the lowest level tested. Although this was not a direct test of the efficacy of our experimental stimuli to elicit the MER, the levels of the experimental stimuli were much weaker than those used to test for the presence of the MER (see below).

All subjects were tested for the presence and strength of spontaneous otoacoustic emissions (SOAEs). Specifically, it was of interest to determine whether any of the subjects had strong SOAEs near the frequency of the probe tone (4000 Hz) used for the physiological measures reported here, because this might have led to unwanted interactions. No subject had an SOAE stronger than -15.0 dB within 600 Hz of the probe-tone frequency; one subject had an SOAE within 200 Hz of the probe tone,

but the magnitude of this SOAE was weak (3938 Hz, -21.3 dB). Therefore, we did not anticipate significant interactions between our subjects' SOAEs, and the elicited nSFOAE responses.

The behavioral measures will be described first, followed by the physiological measures, and then a description will be provided of how the two types of measures were integrated in the experimental paradigm.

METHODS: BEHAVIORAL MEASURES

Each subject was tested individually while seated in a reclining chair inside a double-walled, sound-attenuated room. Two insert earphone systems delivered sound directly to the external ear canals. (The earphone systems will be described in detail in the physiological methods section below.) A computer screen attached to an articulating mounting arm was positioned by the subject at a comfortable viewing distance, and was used to provide task instructions, and trial-by-trial feedback. A keypad was provided to the subject to indicate his or her responses.

Selective Auditory-Attention Conditions

There were two selective auditory-attention conditions: one required dichotic listening, and one required 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 that received the female talker was selected trial by trial from a closed set of random permutations. The number of trials having the female voice presented to the right ear was approximately equal to the number of trials having the female voice presented to the left ear. On each trial, the two talkers simultaneously said 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. Consecutive digits were separated by a 300-ms interstimulus interval (ISI). As described below, the stimulus waveforms used to elicit the nSFOAE response were presented in the ISIs between digits.

28

Figure 1 shows an example of the speech waveforms presented on a single trial of the dichotic-listening condition. The subject always was instructed to attend to the ear in which the female was talking, to recall the seven-digit sequence that she said, and then to match a subset of that sequence to one of two choices presented visually on a computer screen at the end of the trial. Each choice consisted of five digits, and the correct choice always corresponded to the middle five digits spoken by the female talker. The incorrect choice differed from the correct choice by only one mismatched digit, and it was unrelated to the digit sequence spoken by the male talker.

A 2000-ms silent response interval occurred at the end of each trial. The subject responded by pressing one of two keys on a keypad ("4" or "6") to indicate that the correct series of digits was displayed on the left or the right side of the computer screen, respectively. (The "5" key had a raised nipple so that the subject knew where his or her fingers were located without having to look at the keypad.) Immediately following the behavioral response, the series of digits selected by the subject was indicated on the screen with an illuminated border, and feedback was provided to the subject in the form of an illuminated LED above the correct choice.

The diotic-listening condition was similar to the dichotic-listening condition with the exception that the male and female talkers were presented simultaneously to both ears on each trial, rather than to separate ears on each trial. Thus, the dichotic-listening condition required attention to one of two spatial locations, namely 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. Figure 2 shows an example of the stimulus waveforms presented on a single trial of the diotic-listening condition.

Across all subjects and conditions, there were at least 30 behavioral trials per block of trials. However, as described below, the time required to complete a block of behavioral trials was based upon the rate of acquisition of the physiological measure, the nSFOAE, rather than the subject's behavioral performance. That is, when the physiological response from a trial was judged to be unacceptable (see below), an additional trial was added to the block. This was true for the experimental and control conditions alike. In terms of the data, this meant that the behavioral measures such as

percent correct (P(c)) performance and mean reaction time were based upon slightly different numbers of trials for the different subjects in each of the different conditions. The average duration of a block of trials was about 4 – 6 min.

Percent-Correct Performance

For each block of trials, the percentage of correct responses was calculated automatically, and saved to disk. Across repeated measures of a condition, the percentages from four to six blocks of trials were averaged to estimate overall task performance, and this was true for both the dichotic- and diotic-listening conditions. Achieving a certain standard of behavioral performance in the selective-attention conditions was not a primary consideration of this experiment, but rather the more important goal was to design behavioral tasks that were cognitively challenging, and that required the subject to maintain an attentive state in order to perform well. The selective-attention conditions proved to be moderately difficult for most of the subjects, but the number of correct trials always exceeded the number of incorrect trials. Responses in the inattentive-listening condition (see below) were not evaluated in terms of correctness, only in terms of occurrence.

Reaction Time

A reaction time (RT) was calculated for each behavioral response, corresponding to the time that had elapsed between the end of the last 500-ms spoken-digit epoch in a series and the time at which the subject indicated his or her decision with a key press. The maximum reaction time (the duration of the response interval) was 2000 ms; if a behavioral response was not recorded during the designated response interval, neither the behavioral data nor the physiological data from that trial were saved. No-response trials occurred infrequently. Reaction times were recorded and averaged separately for correct and incorrect trials.

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

At the start of an experimental condition, the twenty digital waveforms (ten from the female talker, 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 were equalized in level using the following procedure. The mean rms of the ten digits spoken by the female talker was calculated, and then each of the twenty waveforms was scaled by a fixed amount such that the overall level of each waveform was about 50 dB SPL.

Subject Training

As mentioned previously, each subject was screened for normal hearing prior to participating in the experiments described here. During that same session, a subject who passed the screening was given training with the simultaneous behavioral and physiological procedures by completing several blocks of trials of the inattentive-listening condition (see below). The subjects were not given training with the selective auditory-attention conditions prior to data collection. However, the two male subjects (L01 and L03) 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.

METHODS: PHYSIOLOGICAL MEASURES

Each subject was tested individually while seated comfortably in a double-walled, sound-attenuated room. The nSFOAE-eliciting stimuli were delivered directly to the ears by two insert earphone systems. For the left ear, one Etymotic ER-2 earphone (Etymotic, Elk Grove Village, IL) presented the stimuli through a plastic sound-delivery tube that terminated at a foam ear-tip that was fitted in the external ear canal. Because no microphone was present in the left ear, nSFOAEs were not recorded there. For the right ear, two Etymotic ER-2 earphones 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. nSFOAE responses were elicited by sounds presented by both ER-2 earphones (see below), and were recorded using the ER-10A microphone. nSFOAEs were recorded from the right ear only, even though the nSFOAE stimuli were presented to both ears simultaneously.

The stimulus waveforms (speech and nSFOAE) and the nSFOAE responses both 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 pre-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 – 15.0 kHz), before being passed to the analog-to-digital converter in the sound board.

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). The SFOAE is an OAE that is elicited from the cochlea by a long-duration tone,

recorded in the external ear canal, and analyzed at the same frequency as the eliciting tone (Kemp, 1980). The nSFOAE simply is one version of the SFOAE.

Our nSFOAE procedure is based upon the "double-evoked" procedure described by Keefe (1998). A single estimate of the nSFOAE response was obtained using a cancellation procedure that was performed using three consecutive SFOAE measures (recordings), and these three physiological measures will be referred to as a "triplet." The three measures of a triplet each consisted of the sound recorded in the external ear canal during the presentation of the SFOAE-eliciting stimulus. For the first SFOAE measure of a triplet, the eliciting stimulus was presented using one of the two ER-2 earphones in the right ear. For the second SFOAE measure of a triplet, the same stimulus was presented using the other earphone. Finally, for the third SFOAE measure of a triplet, both earphones presented that same stimulus simultaneously, and in phase. Because the two earphones were activated at the same time for the third measure, and calibrated to the same level, the stimulus waveforms summed in the ear canal, and the resultant stimulus reaching the tympanic membrane was approximately double the amplitude of, and thus approximately 6 dB SPL stronger in magnitude than, the stimulus presented by either of the two earphones alone.

Each recording from the external ear canal necessarily contained both the SFOAE-eliciting stimulus and the acoustic response from the cochlea (the SFOAE), plus any extraneous sounds caused by subject movement, physiological noise, or ambient noise. In order to extract the nSFOAE, the first and second SFOAE responses of a triplet were summed, and the third response was subtracted from this sum. If only linear processes were operating to produce the individual SFOAE responses contained in each of the three recordings, and if the recording system itself was linear, the result of this subtraction would have been a waveform whose magnitude was not discriminable from that of the noise floor of our measurement system. However, the result always was a non-zero waveform representing the amount by which the amplitude of the sum of the first two triplet recordings *exceeded* the amplitude of the third triplet recording, and whose magnitude greatly exceeded that of the noise floor. A stable estimate of the nSFOAE was achieved by averaging the difference waveforms across many Triplets in the same block of trials.

The waveforms obtained from the nSFOAE cancellation procedure were thought to be attributable primarily to differences in the amplitudes of the cochlear responses across triplet presentations, namely because the stimulus waveforms were assumed to have canceled. In support of this assumption, when the same procedure was used with a passive cavity (a 0.5-cc syringe) instead of a human ear, nSFOAE magnitudes were approximately -10 to -15 dB SPL, values commensurate with other estimates of the noise floor of our OAE-measurement system, as measured during experimental data collection. Additionally, because two earphones were used (each presenting the same sound twice)—rather than using the same earphone for all triplet presentations—any nonlinearities introduced by (inherent in the physics of) the earphones themselves also were expected to cancel. Finally, because the slope of the input / output function of the healthy cochlea is compressive (nonlinear) over the range of moderate sound-pressure levels (Bacon, Fay, & Popper, 2004), the cochlear response to the third triplet presentation (the two-earphone presentation) was expected to be less than double the amplitude of the response to either the first or the second triplet presentation (the single-earphone presentations).

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 signal-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. (The use of a single sample of frozen noise was thought to be important for the purpose of averaging nSFOAE waveforms across repeated conditions, and for comparing these averaged responses across the various experimental conditions, and this proved to be the case.)

Figure 3 shows the stimulus waveforms presented during a single trial of the nSFOAE procedure. The top trace shows the waveforms presented to one ear, and the bottom trace shows the waveforms presented to the other ear; the two traces are

identical. On every trial, two triplets were presented consecutively to both ears; thus, the third and sixth stimulus waveforms are shown as twice the amplitude of the remaining stimuli in the series. The nSFOAE responses from Triplet 1 always were averaged and analyzed separately from the responses from Triplet 2. Although not shown in Figure 3, six 30-ms recordings in the quiet immediately followed each of the six SFOAE recordings in the two triplets, and were used to estimate the average (nonlinear) magnitude of the physiological noise floor during each triplet of each behavioral trial (this proved to be an important control; see below for further detail). A nominally 500-ms ISI was placed between consecutive nSFOAE stimuli, and also at the beginning and the end of each stimulus series. As described above, the seven speech stimuli used for the behavioral task were presented during these intervals. In fact, the ISI was selected randomly from the range of values between 500 – 520 ms in order to reduce the periodicity produced by repeated presentations of the stimulus series. On every trial, a 2000-ms silent response interval followed the last stimulus of Triplet 2. Thus, a single trial of the nSFOAE procedure lasted approximately 7 s. Recall that the on every trial, the male and female digit sequences were different (but simultaneous) in the two ears, that the nSFOAE stimulus series was the same (and simultaneous) in the two ears, and that cochlear responses were recorded only from the right ear.

The nSFOAE procedure began with a calibration routine, during which no speech stimuli were presented. For the first 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 factor then was used to scale the amplitude of the experimental stimulus used to elicit the nSFOAE (the tone-plus-noise stimulus described above). This calibration routine was followed by two criterion-setting routines, and the main data-acquisition routine. The first criterion-setting routine consisted of 12 trials (each having two Triplets) during which all nSFOAE responses were accepted 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 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. The median rms voltage from this recording was calculated, and was used as a measure of the ambient (physiological) noise level of an individual subject.

During the main acquisition routine, each new nSFOAE response was compared to the data collected during the two criterion-setting routines, and was accepted into the accumulating nSFOAE average if either one of 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. Recall that 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. The block of trials terminated when each of the two averages was composed of at least 30 nSFOAEs. This typically occurred in about 35 – 50 trials, or about 4 – 6 minutes into a block of trials.

METHODS: BEHAVIOR AND PHYSIOLOGY

As noted above, the stimuli used to elicit the nSFOAE responses were interleaved with the speech stimuli used for the dichotic- and diotic-listening conditions. During each experimental session, in addition to these selective-attention conditions, nSFOAE responses also were measured during an inattentive-listening (control) condition. This control condition was designed to be comparable to the experimental conditions with the exception that it did not require the attention of the subject, or at

least not to the same extent (there was no speech presented, nor was there a forced-choice task on every trial). To explain how the behavioral and physiological procedures were implemented simultaneously, the dichotic-listening condition will be described in detail first, followed by a description of the inattentive-listening condition.

Figure 4 shows an example of the acoustic waveforms presented to the ears during a single trial of the dichotic-listening condition. This figure is an overlay of the behavioral stimuli shown in Figure 1, and the physiological stimuli shown in Figure 3, and illustrates that the speech stimuli (about 50 dB SPL each) were weak relative to the nSFOAE stimuli (about 60 dB SPL each). (The same can be imagined for the diotic-listening condition, if Figures 2 and 3 were overlaid.) As in previous figures, the trace at the top of this figure contains the attended speech series (female talker) plus the interleaved nSFOAE stimuli, and the bottom trace contains the unattended speech series (male talker) plus the interleaved nSFOAE stimuli. The two triplets per trial are marked. Although nSFOAE responses were measured from the right ear only, the same nSFOAE stimulus series also was presented to the left ear on every trial. This was done to ensure that task difficulty was equated across the ears, irrespective of whether physiological responses actually were measured from that ear. Recall that on one half of the trials the subject listened to the female voice in the left ear, from which nSFOAEs were not measured.

As described above, the subject received feedback on the computer screen at the end of every trial as to whether his or her behavioral response was correct or incorrect. In addition, feedback was provided on every trial (also on the computer screen) as to whether the nSFOAE response(s) from one or both (or neither) of the two Triplets was accepted for averaging. In this way, the subject had the potential to recognize coincident occurrences of increased physiological noise and decreased acceptance of the nSFOAE responses, and to change his or her behavior accordingly. Note, however, that the nSFOAE responses were evaluated for acceptance into the accumulating nSFOAE average *only* if a button-press response was recorded on the corresponding behavioral trial. The physiological data from no-response trials were not retained.

37

For the dichotic-listening condition, the nSFOAE responses were averaged separately from trials during which the subject attended to the voice in the right (ipsilateral) ear, and from trials during which the subject attended to the voice in the left (contralateral) ear. (The terms ipsilateral and contralateral are in reference to the right ear of the subject—the ear from which the nSFOAEs always were measured—not the ear that received the female voice.) In this way, nSFOAE responses could be compared from the same subject and ear, measured during the same, brief acquisition period, with the only distinction having been the "direction" of the subject's attention (i.e., "where" he or she was listening). For the diotic-listening condition and for the inattentive-listening condition (see below), there was no such distinction across trials regarding the subject's "direction" of attention.

The nSFOAEs measured during those trials having correct behavioral responses were analyzed separately from the nSFOAEs measured during those trials having incorrect responses. The motivation for this division of the data was to examine whether poorer behavioral performance was associated with a degradation of the physiological responses from the cochlea. One caveat, however, is that the number of incorrect trials always was many fewer than the number of correct trials. Thus, the averaged nSFOAE responses from incorrect trials were less robust than the averaged responses from correct trials.

Inattentive-Listening Condition

A control, inattentive-listening condition was used for comparison with the two selective auditory-attention conditions described above. The rationale for this control condition was to assess whether the cochlear responses recorded during this behavioral task (see below), which did not require the subject to attend to the sounds presented to the ears, might differ from the cochlear responses recorded during tasks that required selective auditory attention. During each trial of the inattentive-listening condition, a series of non-speech stimulus waveforms was presented to the two ears instead of spoken digits. The stimulus series consisted of the six waveforms used to elicit the nSFOAE responses, plus seven 500-ms bursts of speech-shaped wideband noise (the same interleaved arrangement as in Figure 4). Figure 5 is a schematic of the

inattentive-listening condition. Notice that the envelopes of the speech-shaped-noise bursts in this figure match the envelopes of the speech waveforms shown in Figure 1, because they were in fact the same.

The speech-shaped noise bursts were constructed by taking the Fast Fourier Transform (FFT) of each of the 20 speech waveforms used in the auditory-attention conditions and creating 20 wideband noise bursts having those same frequency and amplitude spectra. Then, the temporal envelope was extracted from each speech waveform using a Hilbert transformation, and was applied to each corresponding noise burst. Finally, the envelopes were lowpass filtered at 500 Hz to limit any moment-to-moment fluctuations in the amplitudes of the waveforms. The resultant waveforms were not intelligible as speech. Similar to the dichotic-listening condition, one ear received a series of noise bursts shaped by the speech envelopes from the female talker, and the other ear received a series of noise bursts shaped by the speech envelopes from the male talker. Different series of speech-shaped noises were presented on different trials, having been selected randomly with replacement. The ear that received the "female" noise bursts was selected randomly on each trial from a random set of permutations that equated the number of trials of that block during which the "female" noise bursts were presented to the right versus left ear. The task of the subject simply was to press the number "4" on the response keypad at the end of each trial, after the final sound in the stimulus series.

The inattentive-listening condition was designed to be comparable to the two selective-listening conditions in terms of the stimulus waveforms presented to the ears, and to differ only in the amount of cognitive resources required to perform the behavioral task. The dichotic-listening and diotic-listening conditions had both selective-attention and memory components, but the inattentive-listening condition had neither of these components. Consider for example the dichotic-listening condition: auditory attention was required to locate the female talker at either the left or right ear, and then to maintain focus on the female speech stream while concurrently ignoring the male speech stream. Then, while attending to the female talker, working memory was required to retain the sequence of digits that she spoke until the response interval began. In contrast, in the inattentive-listening condition, the behavioral task only

required that a subject listen for the silent response interval that followed the final sound on every trial.

For the auditory cortex, the sounds presented during the inattentive- and selective-listening conditions differed greatly in terms of their perceptual saliency, but for the cochlea, the sounds presented during every condition should have been processed similarly. For one, the nSFOAE stimuli were identical across all conditions. In addition, the speech-shaped noise bursts differed from the actual speech waveforms only in the amount of perceptual saliency. The spectra were identical, the temporal envelopes were identical, and the overall level of the speech-shaped noise bursts was equal to the overall level of the speech waveforms (about 50 dB SPL). In other words, the speech-shaped noise bursts had the characteristics of speech without actually sounding like speech. Furthermore, the speech-shaped noise bursts were lowpass filtered at 3.0 kHz, just like the actual speech waveforms, and the timing of the presentation of the waveforms was identical in the inattentive- and selective-listening conditions. As in the selective-listening conditions, the physiological responses from the inattentive-listening condition were evaluated for acceptance into the accumulating nSFOAE average only when a behavioral response (a key-press) was recorded. All responses were scored as "correct," and a reaction time was recorded on every trial.

METHODS: DATA ANALYSIS

At the end of a block of trials, the physiological data set consisted of an averaged nSFOAE response from Triplet 1—measured during the first "half" of every behavioral trial—and an averaged nSFOAE response from Triplet 2—measured during the second "half" of every trial. Each of these two averaged responses was composed of at least 30 individual nSFOAE responses. Across two data-collection sessions, each subject completed each condition at least four times, meaning that for each subject there was a set of averaged nSFOAE responses for each of the different listening conditions. Within each set of responses, each averaged nSFOAE response was analyzed offline, and evaluated according to two criteria. Those responses that satisfied both criteria then were pooled to yield a single estimate of the nSFOAE response for that particular condition for that subject. The evaluation and pooling procedures will be described in

detail below. As a result of these procedures, each subject had one pooled nSFOAE response from Triplet 1, and one pooled response from Triplet 2, for each of the listening conditions.[1]

Evaluating nSFOAE Responses

Following data collection, each averaged nSFOAE response (300 ms) was analyzed offline by passing the raw waveform through a succession of 10-ms rectangular 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 – 4.2 kHz (the tonal signal was 4.0 kHz) using a 6th-order, digital elliptical filter. The rms voltage was computed at each 1-ms step, and converted to dB SPL. Figure 6 plots an example of an nSFOAE response from subject L01, averaged from one block of trials. Time is shown along the x-axis, and nSFOAE magnitude is shown along the y-axis. Data were collected with 1-ms resolution, but were plotted at every fifth ms to reduce data density. This will be true for every figure here in which a nSFOAE waveform is plotted.

The first 50 ms of data shows the cochlear response to the 4.0-kHz tone in the quiet, and the last 250 ms of data shows the cochlear response to the tone when presented simultaneously with wideband noise (a dashed, vertical line delineates noise onset). As seen in Figure 6, the response to the tone in the quiet is relatively constant. Following the onset of the noise, the nSFOAE response to the tone shows an immediate decrement in magnitude, followed shortly thereafter (about 25 ms) by an increase in magnitude toward an asymptotic value. The decrement in magnitude was thought to be attributable to mechanical two-tone suppression on the basilar membrane, the more dramatic increase in magnitude (the rising, dynamic segment of the response) was thought to be attributable to neural feedback from the noise-activated efferent

---

[1] For the dichotic-listening condition, the nSFOAE responses from Triplet 1 and Triplet 2 were separated further into ipsilateral and contralateral responses, corresponding to trials during which the subject attended to the female voice in the right or the left ear, respectively. Each of these averaged responses was composed of at least 15 nSFOAE responses (but, as reported in Results, the ipsilateral and contralateral responses later were pooled).

auditory system (Walsh et al., 2010a), and the hesitation seen after noise onset, but prior to the rising response, was thought to be attributable to neural delay (McFadden et al., 2010). The waveform morphology shown here was typical, but there were individual differences in nearly every aspect of the subjects' nSFOAE responses.

All nSFOAE responses were analyzed in the same manner for the purposes of evaluating the individual responses, and for comparing responses across conditions, and across repeated measures of a condition. To simplify the analyses, an estimate of the nSFOAE tone-alone magnitude was calculated by averaging the nSFOAE magnitudes from 5 – 40 ms, using the data at every 1 ms. Similarly, an estimate of the nSFOAE tone-plus-noise magnitude was calculated by averaging the nSFOAE magnitudes from 250 – 285 ms. The data included in these averages are shown as filled circles in Figure 6 (with the caveat that only one fifth of the data used are plotted in the figures). The same number of data points—35: one at every ms—was included in the tone-alone and tone-plus-noise averages. From 60 – 250 ms, shown by the open circles in the figure, nSFOAE responses were fitted with a positive exponential function of the form $y = a (1 - e^{bx}) + c$ (the dotted, curved line in Figure 6) for the purpose of calculating a time constant (τ). If the time constant exceeded the duration of the fitting window (190 ms), the fit was considered to be poor, and that particular nSFOAE response was not assigned a time constant. Using this common temporal window, 250 of 374 total nSFOAE responses—across subjects, triplets, and conditions—were fitted well by the function, leaving about one third of the averaged nSFOAE responses without an estimated time course. Note, however, that there were individual differences in how well the averaged nSFOAE responses were fitted by the exponential function: five of the eight subjects had over 80% of their responses fitted well, with three subjects' responses fit well more than 95% of the time, one subject had 60% of his responses fitted well, and two subjects (L03 and L07) had only 25% or less of their responses fitted well.[2]

---

[2] The total count of nSFOAE responses provided here corresponds to the number of nSFOAEs averaged from the *correct* trials from all conditions, across all eight subjects in the selective auditory-attention conditions. The count of the correct responses was used because the control conditions did not yield an nSFOAE response averaged from "incorrect" trials (all trials were scored as "correct").

Culling nSFOAE Waveforms

For each subject, each set of averaged nSFOAE responses from repeated measures of a listening condition (see middle section of Figure 7) was culled according to two criteria. An nSFOAE was eliminated from further consideration if (1) magnitudes in any part of the waveform were within 3.0 dB of the noise floor (see below) 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. (As explained above, the typical nSFOAE response exhibits a rising, dynamic response following the onset of the wideband noise. The second criterion eliminated nSFOAEs with atypical morphology.) Of the 374 averaged nSFOAE responses collected across subjects, triplets, and conditions, only 18 responses (~5%) were culled according to these two criteria. That is, the vast majority of the data were retained for pooling.

Pooling nSFOAE Waveforms

The culled set of averaged nSFOAE responses from repeated measures of a listening condition was pooled to yield a single estimate of the nSFOAE response for that condition (see rightmost section of Figure 7). The nSFOAEs were pooled independently for Triplet 1 and Triplet 2. Pooling was accomplished by averaging the set of responses point by point after each nSFOAE response had been passed through the analysis window and converted to dB SPL (as shown in Figure 6). The pooling procedure was implemented using custom-written LabVIEW® software.

Physiological Noise-Floor Estimate

For all conditions, a 30-ms recording in the quiet followed the presentation of every nSFOAE-eliciting stimulus. In fact, that silent epoch also was included in every recording of every triplet component involved in the nSFOAE cancellation procedure. In other words, each individual nSFOAE waveform also had a corresponding nonlinear estimate of the physiological noise floor from that particular triplet. (The noise-floor-estimate epochs were not included in the figures above for simplicity of presentation.) The same 10-ms rectangular window that was used to analyze the nSFOAE waveforms was passed through the silent epoch in 1-ms steps. The estimate of the physiological

noise floor began when the onset of the analysis window was at 310 ms—far removed from stimulus offset—and terminated when the onset of the analysis window was at 320 ms—after which point the 10-ms nSFOAE-response analysis window would have extended past the available data. In other words, only the 10-ms segment in the middle of each 30-ms recording was used to estimate the magnitude of the noise floor.

Realistically, there should have been no systematic pattern to the physiological noise estimates across Triplet components, so the result of the nSFOAE cancellation procedure in the silent epoch essentially produced an average estimate of the background noise across a Triplet. The first criterion for culling nSFOAE waveforms eliminated those responses that, at least in part, were not well resolved from the electronic and subject-produced noise that are inherent in any physiological measure. Noise-floor estimates also were compared across conditions in order to examine whether observed differences in nSFOAE magnitudes might be related to systematic differences in the levels of background noise.

Calculating and Interpreting Effect Size

After the nSFOAE data sets were culled and pooled, an effect size (*d*) calculation (Cohen, 1992) was used to compare the pooled responses between the attentive- and inattentive-listening conditions. Tone-alone magnitudes, tone-plus-noise magnitudes, and time constants all were compared in this manner. The formula used to calculate the effect size is shown below (Eq. 1). The numerator in this equation is the difference of the means of the two distributions of data being compared, and the denominator is an estimate of the pooled standard deviation (Eq. 2) across the two distributions.

Eq. 1 $\quad\quad d = (\mu_1 - \mu_2) / \sqrt{\sigma^2_p}$

Eq. 2 $\quad\quad \sqrt{\sigma^2_p} = \sqrt{(((\sigma^2_{p1} * n_1 - 1) + (\sigma^2_{p2} * n_2 - 1)) / (n_1 + n_2 - 2))}$

Effect sizes indicate the separation between the means of two distributions of data in standard-deviation units. By convention, an effect size greater than 0.20 is considered to be small, an effect size greater than 0.50 is considered to be medium, and an effect size greater than 0.80 is considered to be large (Cohen, 1992). For

convenience, this language occasionally will be adopted when describing the calculated effect sizes. For some comparisons, a t-test also was used to evaluate possible differences between sets of data.

Bootstrapping

A bootstrapping technique was used to assess the likelihood of having obtained by chance the effect sizes that were calculated from the experimental data. Sampling distributions of effect size were created by simulating 20,000 replications of the same experiment. In a single replication, the averaged nSFOAE responses from two conditions (e.g., the inattentive- and dichotic-listening conditions) were combined to form a single distribution of data. Then, each individual nSFOAE response was selected randomly without replacement and assigned to one of the two conditions. An effect size was calculated based on the new distributions of data, and stored, and this process was repeated 20,000 times. After the simulations were completed, the result was a sampling distribution of effect sizes against which the experimental effect size could be compared. The implied significance (p) of having obtained a particular experimental effect size was calculated by dividing the number of simulated effect sizes that exceeded the experimental effect by the total number of simulations. In order to provide a relatively conservative assessment of implied significance, the signs of the effect sizes were ignored. Bootstrapping of the experimental data was accomplished using custom-written LabVIEW® software.

Results

BEHAVIORAL RESULTS

Percent-Correct Performance

The selective auditory-attention conditions were designed to require considerable effort on the part of the subjects in order to perform well, but not to be so difficult as to be frustrating, and the percent-correct data suggest that these objectives were satisfied. Subjects performed well above chance in the digit-matching task for both the dichotic- and diotic-listening conditions. Across the eight subjects, the average percent-correct performance in the dichotic-listening condition was 86.2%, and the average percent-correct performance in the diotic-listening condition was 85.9%. Table 2 shows the percent-correct performance for each of the eight subjects in these two conditions. The data in the middle two columns of the table are means calculated across repeated measures of the same condition (on average, each subject completed 4.7 blocks of 30 or more trials per condition), and the data in the bottom row are means calculated across all subjects, and weighted by the number of blocks completed by each subject. The far right column shows the probability that the difference in the performance of an individual subject across the two conditions was due to chance, as calculated using a two-tailed t-test. For this test, the percent-correct data (a non-normal, bounded distribution) were transformed into z-scores (a normal, unbounded distribution). In general, behavioral performance was not statistically different in the two conditions.

For all eight subjects the difference in percent-correct performance between the dichotic- and diotic-listening conditions was small, and was not statistically significant ($p \leq 0.05$). Five of the subjects performed better in the dichotic-listening condition, and three performed better in the diotic-listening condition, revealing no systematic pattern of results. However, for subjects L01 and L05, the diotic-listening condition was consistently, if only slightly, more difficult (smaller percentages of correct trials), and for subject L01 this difference was marginally significant ($p = 0.09$). Subjects L01 and L05 were the only subjects who participated in an earlier, pilot version of this experiment, and thus were the most experienced with these conditions. Their percentages of correct responses were two of the five highest in the group, and their variances were the two smallest.

46

The eight subjects each completed two auditory-attention test sessions. They did not have experience with the selective-listening conditions prior to testing (as described above, subjects L01 and L05 were exceptions). Each condition was run two or three times during each test session, meaning that across-session comparisons involved small amounts of data (mean = 4.7 blocks of trials completed per subject, per condition, across two data-collection sessions). As subjects gained more experience with the selective-listening conditions—from the first test session to the second test session—the percentages of correct performances did *not* improve systematically for the majority of the listeners. (The percentages of correct performance for the dichotic- and diotic-listening conditions were pooled and averaged within a session, and then compared from session one to session two. There were not enough data to have meaningful comparisons across sessions, *within* conditions.) Only one subject demonstrated a statistically significant improvement in performance from the first test session to the second test session, as assessed using a one-tailed t-test. Subject L06 scored 70.8% of the forced-choice decisions correct in session one, and 84.4% correct in session two (p = 0.01).

Reaction Time

Reaction time (RT) was defined as the amount of time elapsed (ms) between the offset of the acoustic stimulus series in a trial and the time at which the subject responded with a key-press. An RT was calculated on every behavioral trial of every condition in this experiment, but only when a behavioral response was recorded within the 2000-ms response interval. If the subject omitted the trial, no RT was recorded. For the dichotic- and diotic-listening conditions, the response interval terminated when the subject pressed one of two keys ("4" or "6") on the keypad to select one of two alternatives in the forced-choice digit-matching task. For the inattentive-listening condition the response interval terminated as soon as the subject pressed the number "4" on the keypad to indicate the end of the final auditory stimulus in the series. Table 3 shows the mean RTs for the eight subjects in the inattentive-listening condition, and in the dichotic- and diotic-listening conditions. Each subject has three rows of data: RTs calculated across all trials of a condition are shown in the first row, and RTs calculated

47

for correct and incorrect trials independently are shown in the next two rows, respectively. The inattentive-listening data were not sorted by trial outcome because there was no judgment of correctness of the behavioral responses (recall, however, that for all conditions, any given physiological response was assessed for inclusion in the nSFOAE average only if a behavioral response was recorded during the designated response interval).

Not surprisingly, reactions times generally were fastest in the inattentive-listening condition, during which no forced-choice decisions were required based on the sounds presented to the ears. The results of two-tailed paired t-tests indicated that, across all eight subjects, the RTs in the inattentive condition were significantly faster ($p \leq 0.025$, adjusted for multiple comparisons) than RTs in both the dichotic- and diotic-listening conditions ($p = 0.001$ for both). For both of the selective-listening conditions, RTs were slower on average for incorrect trials than for correct trials, and this was true for all eight subjects. A two-tailed t-test indicated that the average RT for correct trials was significantly faster than the average reaction time for incorrect trials, and this was true for both the dichotic-listening ($p = 0.00003$) and the diotic-listening conditions ($p = 0.002$).

Across the eight subjects, reaction times on correct trials tended to be slightly faster on average in the diotic-listening condition as compared to the dichotic-listening condition, but this difference was not statistically significant, as indicated by the results of a two-tailed t-test ($p = 0.85$). When reaction times were analyzed separately for correct and incorrect trials, there was no significant difference in reaction times for correct trials between the two conditions ($p = 0.91$), but the difference for incorrect trials was marginally significant ($p = 0.09$). For six of the eight subjects, and in the averages across subjects, reaction times from incorrect trials were faster in the diotic-listening condition than in the dichotic-listening condition, but note that the number of incorrect trials (and thus the RTs) corresponding to these two conditions was relatively small. Behavioral comparisons such as these were not the primary interest, but taken together, the percent-correct-performance and reaction-time data indicated that the subjects were actively engaged in the tasks.

PHYSIOLOGICAL RESULTS

Dichotic Listening: Ipsilateral v. Contralateral Responses

In the dichotic-listening condition, the task of the subject was to attend to the female talker in either the right (ipsilateral) or left (contralateral) ear, and the ear to which the subject's attention was directed was determined randomly on each trial. The primary question addressed *within* this condition was whether cochlear responses in the ear from which the nSFOAEs were recorded—always the right ear—would depend upon the "direction" of a listener's attention. Said differently, the question was whether an individual ear would respond differentially when it was attended versus when it was purposefully ignored. The results indicated that for the majority of subjects, the cochlear response from the right, ipsilateral ear did not differ when subjects attended to the right versus left ear.

Figure 8 shows the averaged nSFOAE responses recorded from correct trials during the dichotic-listening task for each of the eight subjects.[3] Each horizontal pair of panels contains the data of an individual subject. The panels on the left side of the figure contain the nSFOAE responses recorded during Triplet 1, and the panels on right side of the figure contain the nSFOAE responses recorded during Triplet 2. Each panel contains both the averaged ipsilateral (right ear attended) and contralateral (left ear attended) responses of an individual subject. Within a single panel, the nSFOAE response recorded during trials when the right ear was attended is indicated by the solid line, and the nSFOAE response recorded when the left ear was attended (and the right ear ignored) is indicated by the dashed line. Note that the range of nSFOAE magnitudes is the same for all panels, but that the scale for subject L03 has been shifted up by 5.0 dB to better capture the data. For all nSFOAE waveforms plotted here, the first 50 ms of the response was elicited by a 4-kHz tone in the quiet, and the final 250 ms of the response was elicited by that same tone presented in a frozen wideband noise (see Figure 6 in Methods). In general, the effect of the noise was to

---

[3] All of the nSFOAE data presented in the Results were averaged only from trials during which subjects made a correct behavioral decision. The nSFOAE data averaged from incorrect trials had been intended for use in a planned comparison with the data from correct trials, but the incorrect trials proved to be insufficient in number.

increase the magnitude of the nSFOAE response to the 4-kHz tone—the "rising, dynamic segment" described in the Methods.

The data in Figure 8 illustrate several general properties of the nSFOAE response, and these properties are common to every nSFOAE response collected across subjects and conditions. First, as described in the Methods section, the nSFOAE response to a tone presented in the quiet had a relatively constant magnitude, and the nSFOAE response to a tone presented in wideband noise typically increased in magnitude, and then reached an asymptotic value after a few hundred ms had elapsed since noise onset (the rising, dynamic segment). Second, the largest change in the nSFOAE response from Triplet 1 to Triplet 2 was a decrement in the magnitude of the tone-alone portion of the response. In the dichotic-listening condition, for example, tone-alone magnitudes decreased from Triplet 1 to Triplet 2 by 5.6 dB, on average. There typically was much less change in the tone-plus-noise magnitudes from Triplet 1 to Triplet 2 (0.1 dB, on average). Third, and related directly to the second point, the change in nSFOAE-response magnitude from the tone-alone portion to the tone-plus-noise portion typically was larger in Triplet 2 than in Triplet 1. In the dichotic-listening condition, for example, the average change in the magnitude of the nSFOAE response was 10.5 dB in Triplet 2, and 5.1 dB in Triplet 1—about twice as large in Triplet 2. Finally, notice that the moment to moment fluctuations in the tone-plus-noise nSFOAE responses were very regular, particularly within subjects. This observation is a direct reflection of the use of a non-random, frozen-noise sample.

The data plotted in Figure 8 also illustrate that there were notable individual differences in the morphologies of the eight subjects' nSFOAE responses. For example, subjects L03, L05, and L07 exhibited such large tone-alone magnitudes in Triplet 1 that the rising, dynamic segment of the nSFOAE response seen in the other subjects' data largely was missing. For all three of these subjects, the difference in magnitudes between the tone-alone and tone-plus-noise portions of the nSFOAE responses increased in Triplet 2 (thus, a larger change in the nSFOAE response was observed). A second individual difference that can be observed in Figure 8 is the amount of change in the nSFOAE response from Triplet 1 to Triplet 2. Consider the differences in the data between the first four subjects in the top half of Figure 8 (L01,

L02, L03, L04) and the remaining four subjects in the bottom half of the figure (L05, L06, L07, L08). The first four subjects showed very little change in the tone-alone response magnitude from Triplet 1 to Triplet 2 (2.4 dB on average), and the other four subjects showed much smaller tone-alone responses in Triplet 2 (-8.7 dB on average). Across all eight subjects, the tone-plus-noise magnitudes always changed less than the tone-alone magnitudes from Triplet 1 to Triplet 2.

Other individual differences include: (1) the size of the decrement in nSFOAE magnitude immediately following the onset of the wideband noise (thought to be attributable to mechanical, two-tone suppression on the basilar membrane), (2) the duration of this decrement, (3) the size of the increment in nSFOAE magnitude over the duration of the noise stimulus (thought to be attributable to neural feedback from the noise-activated efferent auditory system), and (4) the time course over which the rising, dynamic segments occurred. However, despite these individual differences, for any given subject, the morphology of the nSFOAE response was remarkably consistent (more on this later), thus allowing comparisons *within* subjects to be made with confidence. Individual differences in nSFOAE magnitudes were expected—in light of the outcomes of different nSFOAE conditions run previously that did not require subjects' attention (Walsh et al., 2010a, 2010b)—so most of the comparisons in this experiment were designed to be within subjects, and across the inattentive- and selective-attention conditions. Across-subject data will be discussed only where appropriate.

For all eight subjects the ipsilateral (right ear attended) and contralateral (left ear attended) nSFOAE responses averaged across repeated measures of the dichotic-listening condition were very similar. These responses were compared statistically using two-tailed t-tests, and this comparison was done for both the tone-alone (5 – 40 ms) and tone-plus-noise (250 – 285 ms) portions of the nSFOAE responses, and for the responses obtained from both Triplet 1 and Triplet 2, as described in the Methods. For six of the eight subjects, there was no statistically significant difference between the magnitude of the ipsilateral and contralateral response in either the tone-alone or tone-plus-noise portion of the nSFOAE waveform, and this was true for both Triplet 1 and Triplet 2. In other words, the ipsilateral and contralateral data, which were collected

from the same ear, during the same block of trials, appeared to have originated from the same distribution of cochlear responses in the majority of our subjects.

There were two exceptions to this similarity between the ipsilateral and contralateral responses: subject L03 showed a statistically significant difference between the tone-plus-noise portions of the ipsilateral and contralateral responses in Triplet 2 (p = 0.052), and subject L05 showed a statistically significant difference between the tone-alone portions of the ipsilateral and contralateral responses in Triplet 2 (p = 0.047). For subject L03, the magnitude of the average ipsilateral response was 1.6 dB greater than the magnitude of the average contralateral response, and for subject L05, the magnitude of the average ipsilateral response was 2.8 dB greater than the magnitude of the average contralateral response. Subject L03 completed the dichotic-listening condition four times, and all four times the average magnitude of the ipsilateral response was greater than the average magnitude of the contralateral response in the tone-plus-noise portion of the response for Triplet 2. Similarly, subject L05 completed the dichotic-listening condition six times, and every time the average magnitude of the ipsilateral response was greater than the average magnitude of the contralateral response in the tone-alone portion of the response for Triplet 2. These data suggest that, for these two subjects, attending to the right ear did indeed result in larger nSFOAE magnitudes being elicited from the right ear than attending to the left ear, albeit in different portions of the nSFOAE waveforms for the two subjects.[4]

The dichotic-listening data for subjects L03 and L05 suggest that there may be differences in the extent to which the "direction" of a subject's attention can affect the physiological responses of the cochlea. Individual differences of this sort might manifest themselves in ways in addition to nSFOAE magnitudes, such as in the phases

---

[4] Subject L05 also participated in a version of this experiment that was run previously, and that used the same dichotic-listening condition, but a slightly different control condition. In that experiment, the pattern of the data for this subject in the dichotic-listening condition was the same as shown here: the magnitude of the tone-alone portion of the nSFOAE waveform was greater in Triplet 2 when the right (ipsilateral) ear was attended versus when it was ignored. The average magnitude of this difference was 1.9 dB, across repeated measures, but this difference was not statistically significant (p = 0.26).

of the nSFOAE responses, or in changes in the nSFOAE responses over time, or they may be more evident if a linear SFOAE measure were to be used, so subjects of this sort should be singled out for special consideration by future investigators. Furthermore, it is possible that a different selective-listening task—for example, one in which the nSFOAE-eliciting stimulus is the same as what the subjects are attending to in order to perform the behavioral task—would elicit larger changes, or changes in a larger percentage of the subjects. Tasks of this sort have been designed, and proposed for study, and are described in relative detail in the Discussion.

Because there was no statistically significant difference between the ipsilateral and contralateral nSFOAE responses for six of the eight subjects, and in 14 of 16 nSFOAE-response comparisons when combined across Triplets 1 and 2, all ipsilateral and contralateral responses for each subject were pooled to form a single dichotic-listening response—one response from Triplet 1, and one response from Triplet 2. Pooling of the ipsilateral and contralateral responses was achieved by averaging the two averaged nSFOAE waveforms point by point. The ipsilateral and contralateral responses always were composed of an equal number of nSFOAE waveforms, and thus the two distributions of data were weighted equally in the grand averages. The data for subjects L03 and L05 also were pooled, in spite of the significant differences observed in Triplet 2, because the ipsilateral and contralateral responses from the dichotic-listening condition were more similar to one another than to the response from the inattentive-listening condition, and because it simplified the comparisons with the other subjects and conditions.[5]

---

[5] Pooling of the ipsilateral and contralateral responses had the benefit of producing a composite dichotic-listening response that was more comparable to the diotic-listening and inattentive-listening responses in terms of the number of nSFOAE responses that were averaged. Before the pooling, the ipsilateral and contralateral responses had approximately half the number of responses averaged as the other conditions (it was the only condition for which the data were divided into different, hypothesis-driven distributions.) After the pooling, all of the data compared had about the same number of responses pooled per response.

Inattentive Listening v. Dichotic Listening

In the inattentive-listening condition, the task of the subject was simply to press the number "4" on the keypad during the response interval at the end of the final sound of each trial. The subject did not have to attend to the content of the sounds presented to the ears in order to make this response. The stimulus waveforms presented to the ears during the inattentive-listening condition were constructed to be as similar as possible to those used in the dichotic-listening condition: the nSFOAE stimuli were identical, and the speech-shaped-noise bursts substituted for the speech waveforms were composed of identical frequency spectra, had the same temporal envelope, and had the same overall magnitude. The motivation for implementing these parallels in procedure was to be able to interpret the nSFOAE responses across conditions only in terms of the attentional demands imposed upon the subjects, and not in terms of the stimulus waveforms presented to the ears. The results indicated that for five of the eight subjects, the magnitudes of the responses of the cochlea during the inattentive-listening condition were statistically different from the magnitudes of the responses of the cochlea during the dichotic-listening condition. As described below, however, these differences were observed in different portions of the nSFOAE responses across the individual subjects, and moreover the directions of the change differed in different subjects.

Figure 9 plots the average magnitudes of the nSFOAE responses for the inattentive-listening and dichotic-listening conditions. The figure has four panels; in each panel, pairs of bars show the data for the eight individual subjects, and the data averaged across the eight subjects. The left two panels contain the data from Triplet 1, and the right two panels contain the data from Triplet 2. The top two panels contain the data from the tone-alone portions of the nSFOAE responses, and the bottom two panels contain the data from the tone-plus-noise portions of the responses. Data from the inattentive-listening condition are shown as grey bars, data from the dichotic-listening task are shown as striped bars, and error bars show one standard error of the mean. Within each subject's data set, the inattentive-listening data were compared to the dichotic-listening data, and an effect size was calculated. The effect size is shown above each subject's data. These effect sizes were calculated for each comparison by

subtracting the average nSFOAE magnitude obtained from the dichotic-listening condition from the average nSFOAE magnitude obtained from the inattentive-listening condition and dividing by a common standard deviation. In order to test the likelihood of having obtained an effect size of that magnitude or larger by chance, a bootstrapping technique was used, as described in the Methods section (recall that for these counts, the signs of the bootstrapped effect sizes were ignored). The probability (the implied significance) of having obtained each observed effect size was calculated, and these probabilities are shown in tables below. At the far right of each panel, the average magnitude of the nSFOAE responses was calculated across all eight subjects.

The inattentive-listening data, shown as the grey bars in the four panels, illustrate the range of individual differences in nSFOAE magnitudes that were observed across the eight subjects. Tone-alone magnitudes across Triplets 1 and 2 ranged from about -4.0 dB to about 22.0 dB (mean = 8.1 dB, SD = 7.6), and tone-plus-noise magnitudes across Triplets 1 and 2 ranged from about 12.0 to about 23.0 dB (mean = 15.9 dB, SD = 3.6). Notice in Figure 9 that the ranges of nSFOAE magnitudes on the y-axes are the same for the top two panels of the figure as for the bottom two panels, but that in the bottom two panels the scales have been shifted up by 5.0 dB in order to better capture the data.

In general, the eight subjects can be viewed as consisting of two different groups, based upon differences observed in the directions of the experimental effects. For four of the eight subjects (L01, L02, L03 and L04), the effect of selectively attending to the speech sounds presented dichotically to the ears typically was to increase the magnitudes of the nSFOAE responses relative to the responses measured during the inattentive-listening condition. This was true for both the tone-alone and tone-plus-noise portions of the nSFOAE responses, when an experimental effect was observed. However, there were two (of 16 possible) exceptions to this general trend: subjects L01 and L02 both showed smaller magnitudes in the tone-alone portions of nSFOAE responses from the dichotic-listening condition in Triplet 2, relative to the magnitudes from the inattentive-listening condition.

For the other four subjects (L05, L06, L07 and L08), nSFOAE magnitudes typically were smaller during the dichotic-listening condition, and larger during the

inattentive-listening condition. The most obvious violation of this division of the data was the large difference in tone-alone magnitudes in Triplet 1 for subject L05. Other exceptions included the tone-plus-noise data of subject L06 in Triplet 2, and the tone-plus-noise data of subject L05 in Triplet 2. Of these three (of 16 possible) exceptions to this general trend, only the difference in the tone-alone magnitudes for subject L05 was statistically significant (p = 0.0006). As will be elaborated upon at times below, there were other similarities in the data that warranted this grouping of the subjects.[6]

The left side of Figure 9 contains the data from Triplet 1 of the inattentive- and dichotic-listening conditions. For the tone-alone portions of the responses, shown in the top left panel, four subjects (L01, L02, L03, L05) had larger nSFOAE magnitudes during the dichotic-listening condition, three subjects (L06, L07, L08) had larger nSFOAE magnitudes during the inattentive-listening condition, and one subject (L04) showed no difference in tone-alone magnitudes across the two conditions (d = 0.00). For those seven subjects whose nSFOAE responses were affected by the attentional manipulation, three of the effect sizes calculated from their data were "large" ($d \geq 0.80$), three were "medium" ($d \geq 0.50$), and one was "small" ($d \geq 0.20$). For two of the subjects (L05 and L08), the difference between the nSFOAE magnitudes from the inattentive-listening and dichotic-listening conditions was statistically significant ($p < 0.05$), and for one other subject (L07) the difference was marginally significant ($p < 0.10$). Across the eight subjects, there was no difference in the average magnitudes of the tone-alone

---

[6] The two groups of subjects possessed other similarities in their nSFOAE responses that justified this descriptive division of the data. For example, large differences in the amount by which the tone-alone portions of the nSFOAE responses changed from Triplet 1 to Triplet 2 also were observed (compare the changes in tone-alone magnitudes from Triplet 1 to Triplet 2 for the first four subjects at the top of Figure 8, and for the last four subjects at the bottom of Figure 8). A related example concerned differences across conditions and subjects in the amount of change exhibited by the nSFOAE response to the tone owing to the wideband noise (Δ nSFOAE: tone-plus-noise magnitude minus tone-alone magnitude); briefly, subjects L01 – L04 had larger Δ nSFOAE values from the inattentive-listening condition than the dichotic- and diotic-listening conditions, or had no difference between conditions, and subjects L05 – L08 showed the opposite direction of effect. The Δ nSFOAE measure is not described elsewhere in this report. A final example is the fact that the patterns in dichotic-listening data almost always were the same as the patterns observed in the dichotic-listening data, relative to the inattentive-listening data.

nSFOAE responses across the two conditions (d = -0.01), due to the differences across subjects in the directions of the experimental effects.

The bottom left panel of Figure 9 contains the tone-plus-noise data from Triplet 1 of the inattentive- and dichotic-listening conditions. Four subjects (L01, L02, L03, L04) had larger nSFOAE magnitudes during the dichotic-listening condition, four subjects (L05, L06, L07, L08) had larger nSFOAE magnitudes during the inattentive-listening condition, and for three of the eight subjects (L06, L07, and L08) these differences were statistically significant. Four of the effect sizes from these data were large, two were medium, and one was small. For six of the eight subjects the direction of the experimental effect (the sign of the effect size) was the same for the tone-alone and tone-plus-noise portions of the responses from Triplet 1 (excluding subjects L04 and L05). As was true for the tone-alone data from Triplet 1, because the directions of the effects differed across subjects, the average tone-plus-noise response from Triplet 1 from the inattentive-listening condition was not different from the average tone-plus-noise response from the dichotic-listening condition (d = 0.14).

The right side of Figure 9 (Triplet 2 data) is arranged like the left side of the figure (Triplet 1 data), with tone-alone magnitudes plotted in the top panel, and tone-plus-noise magnitudes plotted in the bottom panel. For the tone-alone data, three subjects (L03, L04, and L06) had larger nSFOAE magnitudes during the dichotic-listening condition, and the remaining five subjects showed larger nSFOAE magnitudes during the inattentive-listening condition. However, the differences in magnitudes across conditions were not statistically significant for any of the eight subjects, and were marginally significant only for subject L07. Furthermore, six of the effect sizes were small, one was medium, and one was large. Interestingly, for every subject but one (L04), the effect sizes for the Triplet 2 tone-alone data were smaller (ignoring the signs of the effect sizes) than the effect sizes for the Triplet 1 tone-alone data, and for only three of the eight subjects, the direction of the difference in tone-alone magnitudes was the same from Triplet 1 to Triplet 2.

The tone-plus-noise data from Triplet 2 of the inattentive- and dichotic-listening conditions are shown in the bottom right panel of Figure 9. Five subjects (L01, L02, L03, L04, and L05) had larger nSFOAE magnitudes during the dichotic-listening

condition, and three subjects (L06, L07, L08) had larger nSFOAE magnitudes during the inattentive-listening condition, but for only one subject (L03) was the difference in magnitudes statistically significant (d = 1.62, p = 0.007). There were three large and three small effect sizes calculated from the Triplet 2 tone-plus-noise data, and for the remaining two subjects' data, the differences in magnitudes were essentially zero (see Tables 4 and 5 below for all magnitudes and effect sizes from the inattentive- and dichotic-listening conditions). Similar to the smaller effect-size magnitudes that were observed in the tone-alone data in Triplet 2 versus Triplet 1, the effect sizes for the Triplet 2 tone-plus-noise data were smaller (ignoring sign) than for the Triplet 1 tone-plus-noise data for six of the eight subjects.

Table 4 provides a more complete summary of the Triplet-1 data plotted in the left half of Figure 9. It contains the means and standard errors of the nSFOAE responses for the eight individual subjects from the inattentive- and dichotic-listening conditions. Shown below the data for the inattentive- and dichotic-listening conditions are pooled variances calculated across the two conditions, effect sizes of the differences between the two conditions, and probability estimates of having obtained those effect sizes as assessed using a bootstrapping technique (see Methods). The effect sizes were calculated by subtracting the average magnitude from the dichotic-listening condition from the average magnitude from the inattentive-listening condition. The same measures were calculated across all eight subjects, and are shown in the far right column of the table. Table 4 is arranged like Figure 9 in that the tone-alone data are shown at the top of the table, and the tone-plus-noise data are shown at the bottom of table.

When the data from Triplet 1 were considered as a whole, seven of the 16 effect sizes from the individual subjects' tone-alone and tone-plus-noise data were considered to be large, five of the effect sizes were medium, three were small, and one effect size suggested no meaningful difference in magnitude. Of the seven large effect sizes, five of these represented statistically significant differences in nSFOAE magnitudes between the inattentive- and dichotic-listening conditions, and one additional large effect size was marginally significant. However, none of the five medium effect sizes achieved statistical significance. Evaluation of the data averaged across all eight subjects

revealed large estimates of the variance pooled between conditions, and effect sizes that were close to zero. This pattern also was observed in the across-subject data from Triplet 2, and in the diotic-listening data described in the next section.

Table 5 shows the means and standard errors of the nSFOAE responses from Triplet 2 of the inattentive- and dichotic-listening conditions. As in Table 4, there are pooled variances, effect sizes, and probability estimates provided both below each subject's data, and below the data calculated across all eight subjects. Similarly, the effect sizes were calculated by subtracting the average magnitude from the dichotic-listening condition from the average magnitude from the inattentive-listening condition, and dividing by a common standard deviation. Finally, as before, the tone-alone data are shown at the top of the table, and the tone-plus-noise data are shown at the bottom of table.

When the data from Triplet 2 were considered as a whole, four of the 16 effect sizes from the individual subjects' tone-alone and tone-plus-noise data were considered to be large, one was medium, and the remainder were small or near zero. Of the four large effect sizes, only that for subject L03's tone-plus-noise data was statistically significant (d = 1.62, p = 0.007); on average, the magnitude of her tone-plus-noise nSFOAE response in the dichotic-listening condition was about 3.0 dB larger than her tone-plus-noise magnitude in the inattentive-listening condition. For subject L07, the differences in magnitudes across the inattentive- and dichotic-listening conditions were marginally significant for both the tone-alone (d = 0.93, p = 0.09) and tone-plus-noise (d = 0.89, p = 0.10) portions of the nSFOAE responses. Across all eight subjects, fewer large effect sizes, and fewer statistically meaningful differences were observed for Triplet 2 than for Triplet 1.

As seen in Tables 4 and 5 it was common to observe that nominally medium and large effect sizes (Cohen, 1992) often did not achieve statistical significance. Of the 17 effect sizes across Triplets 1 and 2 that were classified as being medium or large, only 6 comparisons revealed a statistically significant difference from the bootstrapping procedure. Two other effect sizes were marginally significant. Each set of data was composed of a small number of nSFOAE measurements, so this outcome was not surprising, if frustrating.

Inattentive Listening v. Diotic Listening

The inattentive-listening condition that was compared with the dichotic-listening condition also was compared to the diotic-listening condition. In the diotic-listening condition, the male and female talkers were mixed and presented simultaneously to both ears, and so were perceived as originating from a single location in space—the center of the head. This was the main difference between the diotic-listening and dichotic-listening conditions. As was done for the dichotic-listening data presented above, the diotic-listening data were compared to the inattentive-listening data to examine whether nSFOAEs elicited during periods of selective attention would be different from nSFOAEs measured during periods of relative inattention. Recall that the acoustic stimuli used in the inattentive-listening condition were not speech, but were constructed to be as similar to speech as possible.

Figure 10 shows the data from Triplets 1 and 2 of the diotic-listening condition for each individual subject, and the data averaged across all eight subjects. For comparison, the inattentive- and dichotic-listening data from Figure 9 are re-plotted here. The diotic-listening data are shown as black bars, the inattentive-listening data are shown as grey bars (as in Figure 9), and the dichotic-listening data are shown with less contrast to emphasize that here the effect sizes shown above the bars were calculated by comparing the inattentive- and diotic-listening data. The basic arrangement of Figure 10 is the same as in Figure 9: Triplet 1 data are shown in the left half of the figure, Triplet 2 data are shown in the right half of the figure, the tone-alone data are shown in the top two panels, and the tone-plus-noise data are shown in the bottom two panels. In each of the four panels of Figure 10, the data collected from the eight individual subjects are shown as eight groups of three bars each, and the data averaged across the eight subjects are shown as the final group of three bars at the far right of each panel. As before, error bars show one standard error of the mean, and effect sizes and are shown above the data.

Casual examination of Figure 10 reveals that, for every subject, the magnitudes of the averaged nSFOAE responses measured from repeated measures of the dichotic- and diotic-listening conditions typically had the same direction of effect, relative to the magnitudes from the inattentive-listening condition. Said differently, for an individual

subject, nSFOAE magnitudes averaged from the inattentive-listening condition were only rarely intermediate in value to the nSFOAE magnitudes averaged from the two selective-listening conditions. There were four exceptions to this overwhelming trend: L04's Triplet 1 tone-alone data, L02's and L05's Triplet 2 tone-alone data, and L04's Triplet 2 tone-plus-noise data. As a further argument that the diotic-listening data were more similar to the dichotic-listening data than to the inattentive-listening data, for about half of the comparisons (15 of 32) across the eight subjects, two triplets, and two segments (tone-alone and tone-plus-noise) of the nSFOAE response, the difference between the mean of the inattentive-listening data, and the mean of the diotic-listening data was greater than the difference between the corresponding difference between the inattentive- and dichotic-listening data. As for the dichotic-listening comparison, when the data (in each of the four panels of Figure 10) were averaged across subjects, there was essentially no difference in the mean nSFOAE magnitudes measured during the inattentive- and diotic-listening conditions, and this was true for both the tone-alone and tone-plus-noise portions of the responses, and for both Triplet 1 and Triplet 2.

The top left panel of Figure 10 contains the tone-alone data from Triplet 1 of the inattentive- and diotic-listening conditions. Four of the eight subjects (L01, L02, L03, and L05) had larger nSFOAE magnitudes in the diotic-listening condition, four subjects (L06, L07, and L08) had larger nSFOAE magnitudes in the inattentive-listening condition, and one subject (L04) had an effect size close to zero. Importantly, seven of the eight subjects showed the same direction of experimental effect as observed when the Triplet 1 tone-alone data from the inattentive- and dichotic-listening conditions were compared. (The exception was subject L04, who showed no difference between the inattentive- and dichotic-listening data ($d = 0.00$), and essentially no difference between the inattentive- and diotic-listening data ($d = 0.19$).) Five of the non-zero effect sizes were large, and two were small. The difference between the Triplet-1 tone-alone magnitudes from the inattentive-listening and diotic-listening conditions was statistically significant for one subject (L08) ($d = 1.09$, $p = 0.02$), and marginally significant for three other subjects.

The tone-plus-noise data from Triplet 1 of the inattentive- and diotic-listening conditions are shown below the tone-alone data, in the bottom left panel of Figure 10.

Half of the subjects (L01, L02, L03, and L04) had larger nSFOAE magnitudes in the diotic-listening condition, and half of the subjects (L05, L06, L07, and L08) had larger nSFOAE magnitudes in the inattentive-listening condition. This partitioning of the subjects' data was the same as that observed in the Triplet 1 tone-plus-noise data from the inattentive- and dichotic-listening comparison (just as the patterns in the Triplet 1 tone-alone data were the same for both comparisons). More specifically, compared to the inattentive- versus dichotic-listening effect sizes, all eight of the effect sizes calculated from the individual subjects' inattentive- and diotic-listening data had the same sign (i.e., the directions of the experimental effects always were the same for both comparisons). This was taken as evidence that the cochlear responses measured during the two selective-listening conditions were more similar to one another than to the responses measured during the inattentive-listening condition. Six of the eight effect sizes calculated from the inattentive- versus diotic-listening comparison were large, one was medium, and one was small. For two of the subjects (L05 and L06), the differences in magnitude were statistically significant, and for three other subjects (L03, L07, and L08), the differences were marginally significant.

The data from Triplet 2 of the inattentive- and diotic-listening conditions are shown in the right half of Figure 10. For the tone-alone data in the top right panel of the figure, four of the subjects showed larger nSFOAE magnitudes during the diotic-listening condition, and four subjects had larger nSFOAE magnitudes during the inattentive-listening condition. However, just as was true for the Triplet 2 tone-alone data for the inattentive- and dichotic-listening comparison, the differences in magnitudes between the inattentive- and diotic-listening conditions were not statistically significant for any of the eight subjects. Only one effect size was considered large, two effect sizes were medium, one was small, and for the remaining four subjects, the effect sizes from the Triplet-2 tone-alone data were essentially zero. Also similar to the earlier comparison between the inattentive- and dichotic-listening data, all but two subjects had smaller Triplet 2 tone-alone effect sizes (ignoring sign) compared to the Triplet 1 tone-alone effect sizes. For half of the subjects (L01, L03, L04, and L06), the direction of the difference in tone-alone magnitudes was different in Triplet 1 versus Triplet 2. For the remaining four subjects whose direction of experimental effect was the same in Triplets

62

1 and 2, only one subject (L02) had at least small effect sizes calculated in both triplets (in fact, the effect size from Triplet 1 was large, and the effect size from Triplet 2 was medium). Recall that in the inattentive- versus dichotic-listening comparison, the signs of the majority of the effect sizes from the tone-alone data also reversed from Triplet 1 to Triplet 2.

The bottom right panel of Figure 10 contains the Triplet 2 tone-plus-noise data from the inattentive- and diotic-listening conditions. Four of the subjects (L01, L02, L03, and L05) had larger nSFOAE magnitudes in the diotic-listening condition, and the remaining four subjects had larger nSFOAE magnitudes in the inattentive-listening condition. Across all subjects, three of the effect sizes were large, one was medium, two were small, and two effect sizes indicated no difference between conditions. One of the large effect sizes (that for subject L03) was statistically significant ($d = 1.49$, $p = 0.02$), and another (that for subject L02) was marginally significant. For six of the eight subjects, the direction of the difference in magnitudes across the two conditions was the same for the tone-plus-noise data from Triplet 1 as it was for the tone-plus-noise data from Triplet 2. Perhaps the most important point to note, however, was that for all eight subjects, the directions of the experimental effects observed in the inattentive- and diotic-listening comparison were the same as those observed in the inattentive- and dichotic-listening comparison (when at least a small effect size was observed).

A nonparametric sign test was used to test the probability that the observed consistency in the directions of effect within subjects' data sets was due to chance alone. The sign test only required the assumption that, logically, for an individual subject, the direction of the experimental effect observed in the inattentive- and diotic-listening comparison could have been different (opposite) from that observed in the inattentive- and dichotic-listening comparison. In other words, the probability that the two directions of effect (the signs of the effect sizes) would have been the same was $p = 0.50$—the same probability that the directions of effect would have been different. For Triplet 1, across the eight subjects, and for the tone-alone and tone-plus-noise portions of the nSFOAE waveforms, 15 of 16 effect sizes from the inattentive- and diotic-listening comparison had the same sign as the corresponding effect sizes from the inattentive- and dichotic-listening comparison. When all effect sizes were included in the sign test,

the probability that 15 of 16 comparisons would have the same direction of effect due to chance was p = 0.0005. For the sixteenth comparison, both effect sizes were less than d = 0.20, and one of these was d = 0.00 (all other effect sizes d > 0.20). The probability that 15 of 15 comparisons would have the same direction of effect due to chance (i.e., that a fair coin would come up heads on 15 consecutive flips due to chance) was p = 0.00003 (0.50^15).

For Triplet 2, 12 of 16 effect sizes had the same sign in both the inattentive- and dichotic-listening comparison, and in the inattentive- and diotic-listening comparison. When all effect sizes were included in the sign test, the probability that 12 of 16 comparisons would have the same direction of effect due to chance was p = 0.028. The effect sizes from Triplet 2 were smaller on average than those from Triplet 1, and just 10 of 16 comparisons had at least small effect sizes ($d \geq 0.20$) for both comparisons. Of the 10 (essentially non-zero) effects that were considered to be at least "small," 9 effect sizes from the inattentive- and diotic-listening comparison had the same sign as the corresponding effect sizes from the inattentive- and dichotic-listening comparison. The probability that 9 of 10 comparisons would have the same direction of effect due to chance was p = 0.0098. In summary, then, even though the direction of effect between the inattentive- and selective-listening conditions was different across the subjects, it appears very unlikely that the consistencies observed in the nSFOAE-magnitude data within subjects were due to chance alone.

In what follows, Tables 6 and 7 provide the data for the inattentive- versus diotic-listening comparisons from Triplet 1 and Triplet 2, respectively. These tables have the same basic arrangement as Tables 4 and 5 shown previously, and the inattentive-listening data are the same across both pairs of tables. Comparison of Tables 4 and 5 with Tables 6 and 7 highlights the consistencies in the data within subjects just described.

For the inattentive-listening versus diotic-listening comparison, when the data from Triplet 1 were considered as a whole, 11 of the 16 effect sizes from the individual subjects' tone-alone and tone-plus-noise data were considered to be large, one was medium, three were small, and one effect size suggested no difference in magnitude. Of the 11 large effect sizes, three were statistically significant, and six were marginally

significant. In contrast, the data from Triplet 2 show many fewer differences in nSFOAE magnitudes between the inattentive- and dichotic-listening conditions. Across all subjects, and across the tone-alone and tone-plus-noise portions of the responses from Triplet 2, only one of the 16 effect sizes was statistically significant, and one was marginally significant (both effects indicated differences in the tone-plus-noise magnitudes across conditions).

To summarize the results presented thus far, the nSFOAE magnitudes measured during the inattentive-listening condition often were statistically different from the nSFOAE magnitudes measured during the dichotic-and diotic-listening conditions. The directions of the differences, however, were different for different subsets of subjects (half of the subjects typically had larger nSFOAE magnitudes during the inattentive-listening condition, and half had larger magnitudes during the selective-listening conditions). Most of the statistically significant differences, and the largest effects sizes were observed in Triplet 1, not Triplet 2. Furthermore, the dichotic- and diotic-listening data almost always were more similar to one another than to the inattentive-listening data, especially when at least a small experimental effect was observed. A nonparametric sign test revealed that the probability of having observed the same direction of effect in both the inattentive- and dichotic-listening comparison, and in the inattentive- and diotic-listening comparison, was very unlikely to have occurred by chance. In other words, the differences observed within subjects' data sets between the nSFOAE magnitudes from the inattentive- and selective-listening conditions were thought to be attributable to the differential attentional states of the subjects in these two types of listening conditions.[7]

---

[7] If actual speech would have been used in the inattentive-listening condition in lieu of the speech-shaped-noise bursts, it is possible that some of the experimental effects observed here would have been reduced or obscured due to the inherent (bottom-up) saliency of human speech for human listeners. In a pilot experiment, *reversed* speech waveforms were used instead of speech, but these unintelligible waveforms still sounded very much like human speech, and thus were salient for the subjects, and interesting to listen to (as reported by the subjects, and as experienced by the experimenter). Interestingly, nSFOAE magnitudes measured during the reversed-speech inattentive-listening condition typically were intermediate in value to nSFOAE magnitudes measured during the same dichotic-listening condition, and an inattentive-listening condition during which only the nSFOAE-eliciting waveforms were presented to

Although similar in many ways, there also were notable differences in the nSFOAE responses from the two selective-listening conditions. Notice that for the inattentive- versus dichotic-listening comparisons, there were fewer statistically significant and marginally significant effects, and the effect sizes tended to be smaller in general compared to those measured from the inattentive- versus diotic-listening comparison. There were exceptions to this trend, of course, but the majority of the nSFOAE-magnitude data suggested that the diotic-listening condition produced larger experimental effects than the dichotic-listening condition. A simple explanation is that the diotic-listening condition was more difficult.

In the following sections, two other measures from the nSFOAE procedure will be described: time constants fitted to the rising, dynamic segments of the nSFOAE responses, and noise-floor magnitudes that were measured in the ear canal during each trial.

Time Constants

The rising, dynamic segment of each averaged nSFOAE response from a condition was fitted with a positive exponential function for the purpose of calculating a time constant (τ). Across all subjects and conditions, the duration of the fitting window (190 ms) and the range over which the time constant was measured (60 – 250 ms) were the same. If the time constant for a response exceeded the duration of the fitting window (190 ms), the fit was considered to be poor, and the time constant for that response is not reported. In other cases, a reasonable fit to the data was not possible due to the morphology of the averaged nSFOAE response (see the nSFOAE responses from Triplet 1 for subjects L03, L05, and L07 in Figure 8). This was a more common reason for not obtaining an acceptable fit than obtaining a time constant that was unreasonably large.

Table 8 shows the time constants calculated from the averaged nSFOAE responses from the inattentive-, dichotic-, and diotic-listening conditions. The time

---

the two ears. Note, however, that only a couple subjects were tested using this set of conditions, and that the inattentive-listening condition was inadequate due the lack of a motor response like that required in the dichotic-listening condition.

constants for Triplet 1 are shown in the top half of the table, and the time constants for Triplet 2 are shown in the bottom half. Time constants averaged across all eight subjects are shown in the far right column of Table 8. An "X" in place of a time constant indicates that it was not possible to obtain a time constant for one of the two reasons described above. For subject L05, time constants are reported only for Triplet 1, for subject L07, time constants are reported only for Triplet 2, and for subject L03, time constants are not reported because these three subjects had no rising, dynamic segment to some of their nSFOAE responses (see Figure 8).

For Triplet 1, there were patterns in the time constants across conditions for three of the individual subjects. For subjects L01 and L05, the time constants calculated from the dichotic- and diotic-listening conditions were smaller than the time constant from the inattentive-listening condition, and were similar in duration to one another. For subject L04, the opposite direction of effect was observed: the time constant from the inattentive-listening condition was smaller than either of the time constants from the dichotic- and diotic-listening conditions. For the remaining subjects, no clear pattern was evident. Subject L06 did have slightly smaller time constants in the selective-listening conditions, however, suggesting a possible similarity with subjects L01 and L05. When the time constants for Triplet 1 were averaged across subjects, there was no difference across conditions.

Examination of the individual-subject data from Triplet 2 reveals that the time constants were larger in Triplet 2 than in Triplet 1 for every subject whose nSFOAE responses were fitted well by a positive exponential function (unfortunately, for subject L07, there are no time constants from Triplet 1 with which to compare the time constants from Triplet 2). That is, the nSFOAE response typically had a slower time course in Triplet 2 from the tone-alone to the tone-plus noise segments of the response. (Said differently, the differences in magnitude between the tone-alone and tone-plus-noise portions of the nSFOAE responses were greater in Triplet 2 versus Triplet 1.) For five of the six subjects whose nSFOAE responses were fitted well by a positive exponential function, the time constant for the inattentive-listening condition was larger than the time constant for either the dichotic- or diotic-listening condition. Subject L04 was the exception to this trend, but her data were uniquely consistent: in both Triplet 1

and Triplet 2, the time constant calculated from the inattentive-listening nSFOAE response was smaller than the time constants from either of her selective-listening responses.

The data in Table 8 indicated that, for the majority of subjects, the time course over which efferent activation reached its maximum was shorter in the selective-listening conditions versus the inattentive-listening condition. Said differently, it appeared as though selectively attending to the sounds presented to the ears decreased the amount of time it took for the cochlear response to change from its baseline magnitude to the magnitude at which the effects of efferent activation had reached asymptote. Furthermore, across subjects, these differences were more evident, and more pronounced in Triplet 2 than in Triplet 1. This difference is perhaps made more interesting by the fact that the differences in nSFOAE magnitudes across conditions were smaller in Triplet 2 than in Triplet 1.

Physiological Noise-Floor Magnitudes

In previous studies of the effects of attention on auditory peripheral physiology (e.g., de Boer and Thornton, 2007, and Harkrider and Bowers, 2009), one prevalent concern has been the possibility that differential levels of background noise may exist across experimental conditions, due specifically to the added noise that logically could be produced by the motor behavior required to complete an attentional task. These differences in noise levels would have the potential to confound the physiological results, and their interpretation. Cognizant of this concern, we recorded multiple estimates of the physiological noise floor on every trial of every condition in this experiment. As explained in the Methods, a 30-ms quiet interval followed every presentation of the 300-ms nSFOAE-eliciting stimulus (the duration of each recording was 330 ms). The noise-floor estimate began at 310 ms (far removed from stimulus offset), and terminated at 320 ms (beyond which the 10-ms nSFOAE-response analysis window would have extended past the available data). Thus, we obtained ten 10-ms estimates of physiological noise during the short silent intervals that followed every stimulus presentation.

Table 9 shows the average physiological noise-floor magnitudes calculated across repeated measures of the inattentive-, dichotic-, and diotic-listening conditions. Larger negative values indicate a lower noise floor—a quieter recording, on average, in the brief interval following the nSFOAE-eliciting stimulus. Compared to the tone-alone magnitudes of the nSFOAE responses (the smallest nSFOAEs measured), the noise floor always was at least about -10.0 dB lower in magnitude, and on average this difference was about -15.0 dB. Thus, it was safe to conclude that the noise floor had a negligible effect on the nSFOAEs measured in any condition.

Examination of Table 9 reveals a much more interesting pattern in the data that was evident for every subject, and for both Triplet 1 and Triplet 2. Specifically, noise-floor magnitudes always were highest (noisier) in the inattentive-listening condition, and lower (quieter) in the two selective-listening conditions. Furthermore, noise-floor magnitudes in the diotic-listening condition always were lower than in the dichotic-listening condition. In both Triplet 1 and 2, these two patterns were evident in all eight subjects' data. Within each of the three conditions, noise-floor magnitudes were similar across the two triplets. For the inattentive-listening condition, the average noise-floor magnitude across Triplets 1 and 2 was -10.3 dB. Also calculated across both triplets, the average noise-floor magnitude for the dichotic-listening condition was -12.8 dB, and the average noise floor for the diotic-listening condition was -15.7 dB.

In order to estimate the probability that the observed pattern in the physiological noise-floor data occurred due to chance, a nonparametric sign test was used (as was used for the nSFOAE magnitude data above). The probability that the physiological noise-floor magnitudes would have been lower during the two selective-listening conditions compared to the inattentive-listening condition for all eight subjects owing to chance alone is $p = 0.0039$ (0.5^8). This also is the probability that the diotic-listening noise-floor magnitudes would be lower than the dichotic-listening noise-floor magnitudes for all eight subjects, due to chance alone. Thus, these differences in physiological noise-floor magnitude appear worthy of additional future study.

There are at least two possible interpretations of the observations that noise-floor magnitudes were lower in the selective-attention conditions than in the inattentive-listening condition. (A reminder: these physiological noise-floor estimates are nonlinear

measures of magnitude derived from the nSFOAE cancellation procedure, so any interpretation of these noise-floor differences is not simple.) One interpretation was that the subjects were physiologically quieter—cardiac, pulmonary, and musculoskeletal noises were reduced—somehow owing to the differences in attentional demands across the behavioral conditions. However, previous studies suggested that background noise would *increase* during tasks requiring attention, not decrease, presumably due in large part to the motor behavior required of the subject to respond. In this study, subjects always were required to press a button on a keypad after every trial of every condition, so the motor behavior across tasks should have been equated.

The second interpretation assumes that the recording in the quiet that followed each nSFOAE-eliciting stimulus somehow reflects the decaying SFOAE response from the tone presented in noise. The beginning of the noise-floor estimate (310 ms) was far enough removed from the offset of the nSFOAE-eliciting stimulus that contamination of the response by stimulus resonance in the ear canal would not have been a concern, but close enough to the offset of the stimulus that capturing a physiological response now is thought to be possible (a post-hoc thought). (Transient-evoked OAEs (TEOAEs) can be measured 20 ms after stimulus offset, for example.) If this interpretation were accurate, the implication would be that the cochlear response to the 4-kHz tone was reduced in magnitude in the selective-listening conditions versus the inattentive-listening conditions.

Because the noise-floor measures are nonlinear, lower noise floors would have been observed (1) if the magnitudes of the three linear SFOAEs that comprised a triplet all were smaller during selective attention due to increased efferent effects, or (2) if the effect of attention was to linearize further (further because the MOCB does this reflexively) the slope of the input-output function upon which the 4-kHz tone operates. [A more linear underlying input-output function would result in a smaller (nSFOAE) noise-floor estimate because the difference between the sum of the first two recordings of a triplet and the third recording of a triplet would be reduced relative to a more compressive input-output function—see Methods.]

The fact that noise-floor magnitudes always were lower in the diotic-listening condition than in the dichotic-listening condition suggests that, although similar in many

ways, these two selective-listening conditions were different enough—in terms of their cognitive or perceptual demands—to have produced different responses from the cochlea for every subject tested in this experiment. Recall that the two conditions were identical with respect to the working-memory component of the task, but were different with respect to the selective-attention component of the task—where in perceptual space the subjects listened to the female talker. The implication is that the diotic-listening condition was more difficult than the dichotic-listening condition, and, further, that the diotic-listening condition produced relatively larger changes in cochlear responses (relative to the inattentive-listening condition) due to increased activation of the MOCB efferent system. This is an exciting possibility, but again this narrative requires the interpretation that these noise-floor estimates actually were estimates of the physiological responses from the cochlea.

In order to test the possibility that the observed differences in the magnitudes of the noise floors were due to a procedural difference across the three conditions, three blocks of trials from each condition were run, during which the acoustic stimuli were played to and recorded from a syringe (a non-human, passive cavity) whose volume was approximately 0.5 cm^3. These data are shown on the right side of Figure 11. The data on the left side of the figure are the average physical (electrical) noise-floor magnitudes calculated across the eight subjects in the inattentive-, dichotic-, and diotic-listening conditions. For both sets of averages, the noise-floor data from the inattentive-listening condition are shown as grey bars, the data from the dichotic-listening condition are shown as striped bars, and the data from the diotic-listening condition are shown as black bars.

The physical noise-floor magnitudes averaged across the eight subjects showed systematic differences across the three conditions. As described above, and as is evident in Figure 11, the physiological noise-floor magnitudes recorded in the ear canals of our subjects during nSFOAE data collection were lower in the dichotic-listening condition than in the inattentive-listening condition, and lower in the diotic-listening condition than in either of the two other conditions. In contrast, the physical noise-floor magnitudes recorded in the passive cavity were essentially identical across the three conditions. (The larger error bars for the recordings in the syringe are related directly to

71

the fact that only three blocks of trials were run to obtain each of these three averages.) These differences in the two sets of noise-floor measures offered support to the possibility that the noise-floor magnitudes from our subjects reflected differences in the physiological responses of the cochlea across the three conditions.

Discussion

SUMMARY OF RESULTS

1. The selective-listening conditions were similar in difficulty, and were more difficult than the inattentive-listening condition.

Behavioral performance in the selective-listening conditions was not of primary interest in this experiment. Rather, the interest was in designing a behavioral listening task that was difficult enough to require substantial cognitive effort in order to perform better than chance. For all eight subjects, the average percentages of correct performance in the dichotic- and diotic-listening conditions were well above chance (better than 86% in both conditions), and performance in these two conditions was not statistically different. These outcomes suggested that the subjects understood the selective-listening tasks, that the tasks were manageable, and that the dichotic- and diotic-listening conditions were similar in difficulty. Reaction times typically were fastest in the inattentive-listening condition, during which the subjects were not required to attend to the sounds presented to the ears in order to complete the behavioral task—a predetermined key-press—and mean reaction times were not different in the dichotic-versus diotic-listening conditions. This suggested that the cognitively perfunctory key-press required in the inattentive-listening condition was less difficult than the forced-choice decisions required in the selective-listening conditions. For both the dichotic- and diotic-listening conditions, reaction times were slower on average for incorrect trials than correct trials, and this was true for all eight individual subjects.

2. Differences typically were not observed between the ipsilateral and contralateral responses measured within the dichotic-listening condition.

Within the dichotic-listening condition, the nSFOAE responses measured when the subject attended to the female talker in the right, ipsilateral ear were compared to the nSFOAE responses measured when the subjects attended to the female talker in the left, contralateral ear. The responses for both ipsilateral and contralateral trials always were collected only from the right ear (ipsilateral to the OAE microphone), always were collected during the same blocks of trials, and the ear to which the female talker was presented always was determined randomly trial by trial. The results

73

indicated that typically there was no difference in nSFOAE-response magnitude related to whether the subjects attended to the female talker in the right or left ear (i.e., related to the "direction" of the subjects' attention). In light of this outcome, the ipsilateral and contralateral responses were pooled to yield a single composite dichotic-listening response for comparison with the diotic-listening and inattentive-listening responses. Two of the eight subjects did show a statistically significant difference between the ipsilateral- and contralateral-response magnitudes in Triplet 2, and differences like these warrant further investigation.

3. Large experimental and individual differences were evident in the nSFOAE magnitudes across the inattentive- and selective-listening conditions.

When the composite dichotic-listening responses were compared to the nSFOAE responses measured during the inattentive-listening condition, much larger differences in response magnitude were observed across conditions for the majority of subjects (compared to the smaller differences observed *within* the dichotic-listening data set). Half of the eight subjects typically had larger nSFOAE magnitudes in the inattentive-listening condition than in the dichotic-listening condition, and half typically had larger magnitudes in the dichotic-listening condition. Because of individual differences like these, most of the comparisons of nSFOAE magnitude across conditions were made within subjects, not across subjects. Within each subject's data set, the direction of the experimental effect typically was the same for Triplet 1 and Triplet 2, for both the tone-alone and tone-plus-noise responses, but in some cases, particularly for the tone-alone data, the direction of the effect reversed from Triplet 1 to Triplet 2, or the difference observed in Triplet 1 no longer was evident in Triplet 2.

When the inattentive-listening data were compared to the diotic-listening data, similarly unambiguous patterns were evident in the data within subjects, which typically were parallel to the patterns observed in the inattentive- versus dichotic-listening comparison. A greater number subjects showed statistically significant differences in nSFOAE magnitudes in the inattentive- versus diotic-listening comparison than in the inattentive- versus dichotic-listening comparison, however, perhaps indicating differential effects of the two selective-listening conditions on cochlear physiology. With

few exceptions, nSFOAE magnitudes obtained from the inattentive-listening condition were not intermediate in value to the nSFOAE magnitudes obtained from the two selective-listening conditions, and for every individual subject, the direction of the difference in nSFOAE magnitude between the inattentive- and dichotic-listening conditions typically was the same as the direction of the difference in magnitude between the inattentive- and diotic-listening conditions. These observations suggested systematic individual differences in the responses of the cochlea during periods of selective attention versus inattention.

For both the inattentive-listening versus dichotic-listening comparison, and for the inattentive-listening versus diotic-listening comparison, the largest differences in nSFOAE magnitudes typically were observed in Triplet 1. This result could be an indication of the time course of attentional effects on the efferent auditory system, or it could be a reflection of a potential weakness in the experimental design (see below).

4. For most subjects, the time constants indicated that MOCB activation was faster during selective-listening tasks than in the inattentive-listening task.

Within the individual subjects' data sets, the time constants fitted to the rising, dynamic segments of the averaged nSFOAE responses typically were smaller (the time courses were faster) for the dichotic- and diotic-listening conditions compared to the inattentive-listening condition. Said differently, less time was required to reach an asymptotic level of efferent activation in the selective-listening conditions. This effect was particularly evident in the data from Triplet 2, where five of the six subjects (those whose data were fitted well by a time constant) showed this direction of effect, and where this effect also was present in the across-subject averages. The time constants for the two selective-listening conditions typically were similar. One subject (L04) showed the opposite direction of effect for both Triplet 1 and Triplet 2: the time constant from the inattentive-listening condition was larger (the time course was slower) than either of the time constants from the dichotic- or diotic-listening conditions.

5. For all subjects, physiological noise floors were highest during inattentive listening, lower during dichotic listening, and lowest during diotic listening.

A 30-ms quiet interval followed every nSFOAE-eliciting stimulus, and these intervals were included in the nSFOAE cancellation procedure. Physiological noise-floor magnitudes were estimated by stepping the 10-ms analysis window through the middle 10 milliseconds of each interval (from 310 – 320 ms; see Methods for rationale). Thus, on every trial, a nonlinear estimate of the noise floor for Triplet 1, and a corresponding estimate for Triplet 2 were obtained (just as two nonlinear estimates of nSFOAE magnitude were obtained on every trial). Across all eight subjects, and evident in the data for both Triplet 1 and Triplet 2, noise-floor magnitudes always were highest (noisiest) in the inattentive-listening condition, lower (quieter) in the dichotic-listening condition, and lowest (quietest) in the diotic-listening condition. There were no exceptions to this outcome, so this pattern was captured well in the across-subject data.

To test the possibility that the observed differences in noise-floor magnitudes were due to unforeseen procedural differences across the experimental conditions, physical noise-floor magnitudes were recorded with the OAE-microphone assembly placed in a passive cavity (a syringe), and averaged across three blocks each of the inattentive-, dichotic-, and diotic-listening conditions. The data indicated that noise-floor magnitudes were indeed the same across the three conditions when measured in a passive cavity rather than in human ears. The most appealing interpretation of this result was that the noise-floor magnitudes measured in our subjects in fact captured the magnitudes of residual cochlear responses to the nSFOAE stimuli that remained after stimulus offset. If accurate, this interpretation suggests greater inhibition of cochlear responses during selective-listening conditions than during inattentive-listening conditions, and greater inhibition during diotic-listening conditions than during dichotic-listening conditions.

INTERPRETATIONS AND IMPLICATIONS OF RESULTS

We interpret this collection of results as evidence that the average cochlear response to the same acoustic stimulus can differ depending upon the attentional state of the listener. Large individual differences were observed across the eight subjects with respect to the morphologies and magnitudes of the nSFOAE responses obtained, but systematic patterns were present *within* each subject's data set that suggested reliable differences in the magnitudes of cochlear responses when the subjects were engaged with the sounds presented to the ears in order to perform the behavioral task, compared to when the subjects did not have to attend to the content of the sounds presented to the ears. Systematic patterns across the comparisons of nSFOAE magnitude between the two selective-listening conditions (dichotic- and diotic-listening) and the inattentive-listening condition further reinforced the conclusion that the magnitudes of cochlear responses were different in the two selective-listening conditions from responses in the inattentive-listening condition. Unfortunately, the nonlinear nature of the nSFOAE measure makes it difficult to interpret these differences.

More systematic patterns across the eight subjects were observed in the time-constant and physiological noise-floor data. Time constants from the dichotic- and diotic-listening conditions typically were smaller than from the inattentive-listening condition, suggesting that the time course of efferent activation was shorter during selective listening that required focused attention than during inattentive, or more passive listening. In addition, for all eight subjects, noise-floor magnitudes were smaller (quieter) in the two selective-attention conditions, and smallest (quietest) in the diotic-listening condition. This suggested that residual nSFOAE magnitudes measured after stimulus offset reflected differences in the way the nSFOAE stimulus was being processed in the three conditions. In sum, the nSFOAE-magnitude data, the time-constant data, and the noise-floor data provide evidence that cochlear responses change in relation to the perceptual and cognitive demands of the listening situation.

One result (among many) from this experiment that deserves further consideration was that the differences in nSFOAE magnitude between the inattentive-listening and selective-listening conditions typically were larger for Triplet 1 than for

Triplet 2. One reason smaller differences may have been observed in Triplet 2 was that attention was not maintained throughout the entire triplet. Because the final digit in the series was not used in the digit-matching task, it was possible that subjects had finished attending to the information at the ears during Triplet 2, and instead were engaged in rehearsal of the digits, or some other form of retention, before the behavioral response was required.

This set of results potentially has important implications for persons with hearing loss. Listeners with sensorineural cochlear damage suffer from the so-called "double-whammy" of hearing loss: not only are the benefits associated with a healthy cochlear-amplifier system lost or mitigated, such as hearing sensitivity and frequency selectivity, but so too are whatever benefits are conferred from the rich efferent innervation of the cochlear amplifier, such as the improved ability to detect brief signals in noisy backgrounds (Kawase et al., 1993). A better understanding of how the efferent auditory system interacts with the afferent cochlear-amplifier system, and how cognitive processes such as the selective allocation of attentional resources can modify these interactions, may inform the design of products that aid those with hearing loss.

These results also have potential implications for persons with attention-deficit disorders if part of the reason for the difficulties they experience is perceptual in nature. Accordingly, it is of interest to examine whether persons diagnosed with attention-deficit disorders exhibit the same types of differences in their nSFOAE responses—between the inattentive- and selective-listening conditions—that were evident in the subjects' data collected in this experiment.

ADVANTAGES AND DISADVANTAGES OF THE nSFOAE MEASURE

The nSFOAE measure possesses several advantages over other OAE measures that make it attractive for use: (1) The nSFOAE procedure removes the eliciting stimulus from the recording, permitting its use as a perstimulatory measure, meaning that the physiological response elicited from the ear is measured at the same time that the eliciting stimulus is presented to the ear. With other types of OAEs, such as click-evoked OAEs (CEOAEs) and transient / tone-evoked OAEs (TEOAEs), the physiological response only can be viewed after several milliseconds have passed from

the offset of the transient stimulus, and thus the measured response inherently is a dampened version of the original response to that stimulus, and the onset of the response cannot be observed. Using a perstimulatory measure provides the opportunity to use much longer-duration stimuli, and to measure frequency-specific effects over much longer periods of time. In the experiment described here, for example, we measured the cochlear response to a long-duration tone in otherwise quiet, and then observed changes in that response when a long-duration wideband noise was presented simultaneously with the tone. (2) The moderate-level, long-duration tones used to elicit the nSFOAE are less likely than clicks or tone bursts to inadvertently activate the efferent auditory system with repeated presentations (Guinan et al., 2003), and thus are better able to serve as uncontaminated probes of (nonlinear) cochlear functioning, whether or not efferent activation was intended to be part of the experimental design.

(3) The nSFOAE response has a favorable signal-to-noise ratio, and thus corruption of the physiological response by noise rarely is a concern. Across the eight subjects and the three conditions used in this experiment, the average magnitude of the tone-alone response measured from Triplet 2 was 5.3 dB SPL (The tone-alone portion of the nSFOAE response from Triplet 2 typically was the closest portion of any nSFOAE response to the physiological noise floor; tone-alone magnitudes typically were larger in Triplet 1 versus Triplet 2, and the tone-plus-noise magnitudes almost always were larger than the tone-alone magnitudes.) By comparison, the average magnitude of the noise floor measured during the inattentive-listening condition was -10.3 dB SPL, when averaged across all eight subjects' responses measured from both triplets. In other words, the most unfavorable signal-to-noise ratio observed in this experiment was about 15.0 dB.

(4) For an individual subject, the nSFOAE typically is quite stable across repeated measures, both within and across test sessions. Although large estimates of variability were observed in some subjects' data (e.g., subject L01's inattentive-listening data), in general the nSFOAE responses for an individual subject, within a given condition, were remarkably consistent. Consider one example: subject L03 completed the diotic-listening condition three times in the first test session, and three times during

79

the second test session. The Triplet 2 tone-alone magnitudes from the first test session, each from separate blocks of trials, were 21.8, 21.5, and 21.2 dB, and the tone-plus-noise magnitudes were 25.0, 26.6, and 24.0 dB. In the second test session, the Triplet 2 tone-alone magnitudes were 21.5, 22.2, and 21.9 dB, and the tone-plus-noise magnitudes were 22.3, 25.4, and 27.0 dB. Across repeated measures from the two sessions, the standard deviation of the tone-alone magnitudes was 0.4, and the standard deviation of the tone-plus-noise magnitudes was 1.7. For this subject, the consistency was just as remarkable for Triplet 1. One final advantage of the nSFOAE is that (5) we have used this measure successfully in other, previous experiments (Walsh et al., 2010a, 2010b), and thus had certain expectations for the response.

The primary weakness of the nSFOAE is the inherent difficulty in interpreting a physiological response that has been derived from a nonlinear procedure. The nSFOAE is a measure of the deviation from linearity in the magnitude of a response elicited by a stimulus presented at two different levels. Therefore, nSFOAE responses provide an indirect measure of cochlear responses, and nSFOAE-response magnitudes provide only an indirect measure of the linear cochlear-response magnitudes. Because of this fact, informed inferences have to be made regarding the three linear SFOAE responses contained within the three recordings of a triplet that were used for the nonlinear cancellation procedure. (A linear estimate of SFOAE magnitude was (theoretically) measured during each of the three recordings of a triplet, but the relatively weak physiological response was unobservable due to presence of the concurrent and much stronger acoustic stimulus.)

COMPARISONS WITH PREVIOUS REPORTS

As described in the Introduction, comparisons of the experimental results among previous reports is made difficult by significant procedural differences. The present experiment did not mitigate this problem. Rather, the results presented here also are not easily compared to those from previous reports because of the unique procedural details of this experiment. To our knowledge, this is the first experiment (1) to have used a form of long-duration, perstimulatory OAE to examine cochlear responses during attentive and inattentive listening, as opposed to OAEs elicited by transient stimuli like

clicks and tone bursts that have been shown previously to (perhaps inadvertently) activate the MOCB efferent system; (2) to have used cognitively challenging listening tasks—with both selective-attention and working-memory components—during which behavioral performance was evaluated in terms of correctness, consistency, and speed; (3) to have used complex, and inherently salient speech waveforms as the focus of attention to complete the selective-listening tasks, and other, relatively simple waveforms to elicit physiological responses from the cochlea; and (4) to have made behavioral and physiological comparisons between selective-attention conditions that differed in their attentional demands (the dichotic- and diotic-listening conditions) in addition to those comparisons made between the selective-listening and inattentive-listening conditions.

Nevertheless, there is value in attempting to compare the present findings with those reported previously. To that end, consider the following narrative: if the assumption is made that lower physiological noise-floor magnitudes reflected increased MOCB efferent activation (an increase of inhibition of the cochlear-amplifier system), then we could conclude that the least amount of efferent activation was observed during the inattentive-listening condition, that more efferent activation was observed during the dichotic-listening condition, and that the most efferent activation was observed during the diotic-listening condition. In other words, the effect of selectively attending to sound was to increase the amount of inhibitory neural feedback to the cochlea, and this effect was titrated based upon the difficulty of the listening condition. If this interpretation is correct, it lends support to those previous reports that showed a decrease in the magnitudes of cochlear responses during auditory attention (i.e., an *increase* in MOCB efferent activation during auditory attention), such as Froehlich et al. (1990) and Maison et al. (2001), but is in opposition to those reports that showed an increase in cochlear-response magnitudes during auditory attention (an apparent release from efferent inhibition during auditory attention), like Giard et al. (1994) and Harkrider and Bowers (2009).

PRESENT AND FUTURE DIRECTIONS

Linear Measures of OAE Magnitude

As described above, although the nSFOAE measure has certain characteristics that are of experimental value, its primary weakness is that it is more difficult to interpret in terms of magnitude than a linear OAE measure. Using the same inattentive- and selective-listening conditions used in this experiment, we intend to interleave a linear estimate of OAE magnitude—click evoked OAEs (CEOAEs)—with the stimuli used for the behavioral tasks, rather than using the nSFOAE. In lieu of obtaining two estimates of nSFOAE magnitude from every trial, we would thereby obtain (at least) six estimates of linear CEOAE magnitude from those same stimulus intervals (depending upon the exact procedure). The primary advantage of using CEOAEs would be the direct interpretation of the CEOAE magnitudes. Furthermore, the CEOAE data could provide a relatively fine-grain look at the time-course of efferent activation, and of the effects of attention on efferent and cochlear processes.

The most relevant disadvantage of using transient stimuli such as clicks is that they can produce inadvertent activation of the MOCB (as described above), but this concern would be mitigated if the same stimuli were used in all of the listening conditions. Were this control to be implemented, any observed differences in CEOAE magnitudes could be interpreted with respect to differences in the attentional demands across the behavioral conditions. The prospect of using CEOAEs (or another linear OAE measure) has been made more attractive in light of the results from the physiological noise-floor comparisons across conditions in the present experiment. If our interpretation was correct, those noise-floor estimates actually were measures of a decaying physiological response in the absence of an acoustic stimulus – exactly what would be measured were CEOAEs to be used.

Frequency-Specific Listening Conditions

The nSFOAE is a frequency-specific physiological measure when a tone is used as the eliciting stimulus, and when the physiological response is filtered about the frequency of the tonal stimulus. In this experiment, subjects attended to complex speech waveforms that were interleaved with the nSFOAE stimulus series, but we have

designed additional conditions in which subjects attend directly to the same sounds used to elicit nSFOAE responses. Using simple acoustic stimuli like tones as the focus of attention, as opposed to relatively complex speech waveforms, it would be possible to examine the responses of the cochlear-amplifier system during (more traditional) psychophysical tasks, such as those requiring the ability to discriminate differences in the frequencies or levels of consecutively presented sounds, or the ability to localize sounds based upon differences in timing or levels of the sounds presented to the two ears. A perceived advantage of using the same sounds to elicit nSFOAE responses and to guide selective-listening behavior is gaining the ability to know "where" the subject is listening.

These two directions for further research were designed to address some of the following, lingering questions remaining from the literature summarized above:

(1) Can the effects of attention on cochlear processing be measured in the absence of an acoustic stimulus that activates the MOCB? This was the interpretation of several previous studies, including Puel et al. (1988), Froehlich et al. (1993a), Meric and Collet (1994b), and Giard et al. (1994). (2) Are the effects of attention frequency-specific? In other words, does attention to a specific frequency component of an acoustic stimulus change the physiological responses to unattended components in the same stimulus, or in other concurrent sounds? Measuring OAEs within specific spectral regions has the potential to reveal specific innervation patterns of the efferent auditory system. Maison et al. (2001) reported a frequency-specific effect of selective attention—an increase in contralateral suppression of ipsilateral tones when the same frequency was attended to in the contralateral ear. Froehlich et al. (1993a) reported that visual and auditory attention affect different, adjacent regions of the cochlear partition, but frankly it is not obvious what benefit such a scheme would afford. (3) Are unattended sounds within the attended ear affected in the same manner as attended sounds? Do these effects depend on the presence or absence of an acoustic MOCB elicitor?

Finally, we have designed an experiment to address a broader question proposed by the literature: (4) Does visual attention affect cochlear responses? Previous studies have shown what appears to be a consistent effect of visual

attention—an inhibition of cochlear responses to task-irrelevant sounds (see above). Extensive work on this topic has been completed, but only the methodology and some preliminary results are summarized here.

Selective Visual-Attention Conditions

Seven of the eight subjects that participated in the auditory-attention experiment also participated in a comparable visual-attention experiment. There were two visual-attention conditions, which were designed to be similar to the auditory-attention conditions described above, and to differ primarily in the fact that these conditions required subjects to use visual attention rather than auditory attention. Each subject participated in two data-collection sessions, and across those sessions they completed each of the four conditions to be described a minimum of four times. As was true for the auditory-attention sessions, each subject was tested individually in a double-walled, sound-attenuated room. The subject sat in a reclining chair, and two earphone systems delivered sound directly to the external ear canals. A computer screen was positioned in front of the subject, and a keypad was provided to indicate his or her responses.

For both visual-attention conditions, the task of the subject was the same—to attend to one of two series of digits presented visually at adjacent locations in the center of a computer screen. The digits were presented consecutively in black typeface on white rectangular backgrounds with colored borders. The border around one rectangle was pink (artificially "female"), and the other border was blue (artificially "male"). The subject always was instructed to attend to the series of digits that was presented within the rectangle with the pink border. Similar to the auditory-attention conditions, there were seven digits in each series, selected randomly with replacement on every trial, and selected independently for the two series. In addition, the location of the series of digits to be attended (either left or right) was selected on each trial from a closed set of random permutations, and the probability that the to-be-attended location was on the left side of the screen was the same as it being on the right side of the screen.

Figure 12 is a schematic of one trial of the visual-attention condition, proceeding in time from the foreground to the background. Each pair of digits in the two series was presented for 500 ms (to match the acoustic presentation in the previous experiment),

and consecutive pairs of digits were separated by a 300-ms ISI during which no digits were displayed on the screen, only the two white rectangles. Just as in the auditory-attention conditions, the stimuli used to elicit the nSFOAE were presented during the ISIs between digits, and two nSFOAE waveforms were measured on each behavioral trial—one from Triplet 1, and one from Triplet 2. For one of the two visual-attention conditions, the speech stimuli used in the auditory-attention conditions also were presented separately to the two ears (just as in the dichotic-listening condition) at the same time that the digit series were presented on the computer screen. The digits spoken by the female talker were selected independently from the digits spoken by the male talker, and both auditory digit series were selected independently from the two visual digit series, and so provided no information for responding in the behavioral task. The ear to which the female talker was presented was selected randomly on every trial. For the other visual-attention task, the speech stimuli were not presented to the ears, only the stimuli used to elicit the nSFOAE.

Following each trial, two alternative choices were displayed on the computer screen, each composed of five digits. The correct choice always corresponded to the middle five digits in the attended series, and the incorrect choice differed from the correct choice by only one mismatched digit. The subject indicated his or her decision by pressing a key on a keypad. As before, the number "4" corresponded to the choice on the left side of the screen, the number "6" corresponded to the choice on the right side of the screen, and the subject had 2000 ms to indicate his or her decision. Each behavioral response was scored as correct or incorrect, feedback lasting 200 ms was provided on every trial, a reaction time was recorded on every trial, and percent correct performance was calculated at the end of each block of trials. These details matched exactly the behavioral task in the auditory-attention experiment.

Control Conditions

There were two control conditions that also were run during the visual-attention sessions, and that were used for comparison with the two experimental conditions. For both control conditions the task of the subject was to press the number "4" on a keypad at the end of each trial. One control condition was the inattentive-listening condition

described above in the auditory-attention methods, during which speech-shaped noise bursts and the nSFOAE stimuli were presented in an interleaved fashion to the two ears (see Figure 5 in Methods).   In the second control condition, the stimulus series consisted only of the six waveforms used to elicit the nSFOAE responses (as in Figure 3 in Methods).   The nSFOAE data from the inattentive-listening condition were compared to the nSFOAE data from the visual-attention condition with speech, and data from the nSFOAE-only condition were compared to the data from the visual-attention condition without speech.   Within each comparison, the stimulus series presented to the ears were practically identical (the only exception was the difference in the order of presentation of the randomly selected digits and speech-shaped-noise bursts across conditions), and the experimental and control conditions differed only in the amount of attention required to complete the behavioral task.   As was true for the auditory-attention conditions, the principal question addressed by these comparisons was whether the cochlea would process the same sounds differentially depending upon the visual attentional demands imposed upon the subject.

The results from the selective visual-attention conditions will be reported in detail elsewhere.   However, one important outcome was that the physiological noise-floor measurements obtained during the brief silent intervals following each nSFOAE stimulus behaved the same as in the auditory-attention conditions.   Specifically, noise-floor magnitudes were lower in the two visual-attention conditions than in the two control conditions.   This outcome strongly suggests that efferent control of cochlear responses was affected by the attentional demands of a visual task.

Authors	Type of Emission	CL Noise?	Behavioral Task	Focus of Attention	Main Findings
Puel et al. (1988)	CEOAE ("EOAE") dominant freq.	No	Count target letter flashed on screen	Visual	CEOAE amplitude reduced in 13 / 16 subjects
Froehlich et al. (1990)	CEOAE ("EOAE")	Yes	Count target letter flashed on screen	Visual	CEOAE amplitude reduced in 3 / 16 subjects
Froehlich et al. (1993a)	CEOAE ("TEOAE")	No	Reaction time (light flash or tone pip)	Visual Auditory (contra)	CEOAE amplitude reduced during both attention tasks
Meric & Collet (1994b)	SOAE; CEOAE ("TEOAE")	No	Reaction time (light flash)	Visual	CEOAE amplitude reduced in subjects without SOAEs
Giard et al. (1994)	EOAE	No	Attend to tones in one ear	Auditory (ipsi or contra)	EOAE amplitude greater when tones were attended
Maison et al. (2001)	EOAE	Yes	Attend to tones in contra noise	Auditory (contra)	EOAE suppression greater if the same-freq tone attended in CL ear
de Boer & Thornton (2007)	CEOAE	Yes	Arithmetic, or DVD Count pips in ipsi ear	Visual; Auditory (ipsi)	CEOAE amplitude unaffected in all tasks; small effect on I/O slope
Harkrider & Bowers (2009)	CEOAE	Yes	Attend to clicks, or to contra noise	Auditory (ipsi or contra)	CEOAE suppression reduced during both attention tasks

Table 1: Comparison of Previous Results

	Dichotic	Diotic	*p*
L01	86.4	82.8	.09 *
L02	73.0	75.8	.71
L03	95.2	93.0	.29
L04	93.3	90.3	.81
L05	98.5	95.4	.12
L06	81.0	85.8	.34
L07	83.5	86.3	.37
L08	78.6	78.0	.95
AVG	**86.4**	**86.6**	

* Marginally significant difference, two-tailed t-test,
calculated on the z-transforms of the entries shown

Table 2: Percent-Correct Performance in Auditory-Attention Tasks

	Type of Listening Task:	Inattentive	Dichotic	Diotic
L01	All Trials	490.1	1255.5	1240.4
	Correct Trials		1222.4	1198.0
	Incorrect Trials		1473.3	1443.1
L02	All Trials	1487.8	1341.2	1313.4
	Correct Trials		1290.1	1265.2
	Incorrect Trials		1491.0	1474.3
L03	All Trials	360.5	1204.5	1223.5
	Correct Trials		1181.2	1190.4
	Incorrect Trials		1589.4	1672.6
L04	All Trials	228.3	1241.4	1297.5
	Correct Trials		1215.0	1286.0
	Incorrect Trials		1592.4	1407.4
L05	All Trials	955.4	1199.6	1193.1
	Correct Trials		1193.1	1185.2
	Incorrect Trials		1644.2	1354.5
L06	All Trials	996.4	1385.7	1399.1
	Correct Trials		1358.0	1384.2
	Incorrect Trials		1501.2	1484.6
L07	All Trials	404.9	1466.7	1461.9
	Correct Trials		1442.9	1438.8
	Incorrect Trials		1590.1	1608.5
L08	All Trials	772.4	1476.0	1360.5
	Correct Trials		1442.9	1351.5
	Incorrect Trials		1596.3	1392.6
AVG	**All Trials**	**712.0**	**1321.3**	**1311.2**
(n=8)	**Correct Trials**		**1293.2**	**1287.4**
	Incorrect Trials		**1559.7**	**1479.7**

Table 3: Reaction Times in the Auditory-Attention Conditions, by Trial Outcome

Tone Alone	L01	L02	L03	L04	L05	L06	L07	L08	*M*
Inattentive									
Mean	5.31	1.89	21.92	7.21	9.05	7.02	19.62	13.72	*10.72*
Std. Error	1.18	0.77	0.04	0.77	0.60	0.56	0.29	0.70	*2.50*
Dichotic									
Mean	6.69	2.85	22.12	7.22	11.92	5.79	17.77	11.73	*10.76*
Std. Error	0.34	0.58	0.15	0.76	0.37	0.69	0.64	0.24	*2.30*
Pooled Var.	4.62	3.96	0.10	4.67	2.95	3.32	3.76	2.08	*45.86*
Effect Size	-0.64	-0.48	-0.66	0.00	-1.67	0.68	0.95	1.38	*0.01*
p Value	0.23	0.35	0.21	1.00	.0006	0.19	0.08	0.004	
Tone + Noise									
Inattentive									
Mean	14.86	11.78	22.77	11.85	16.19	17.26	21.81	14.92	*16.43*
Std. Error	2.19	0.80	0.59	0.62	0.45	0.49	0.76	0.71	*1.45*
Dichotic									
Mean	17.62	12.74	23.49	13.16	15.34	15.06	18.33	11.27	*15.88*
Std. Error	0.77	0.78	0.67	0.40	0.29	0.53	0.72	0.73	*1.38*
Pooled Var.	17.02	5.71	3.34	2.16	1.72	2.23	5.29	4.44	*15.91*
Effect Size	-0.67	-0.40	-0.39	-0.89	0.65	1.47	1.51	1.73	*0.14*
p Value	0.20	0.43	0.45	0.11	0.13	0.01	0.01	0.005	

Table 4: Average nSFOAE Magnitudes and Effect Sizes from Triplet 1 of the Inattentive- and Dichotic-Listening Conditions

Tone Alone	L01	L02	L03	L04	L05	L06	L07	L08	*M*
Inattentive									
Mean	3.89	0.76	21.72	3.74	2.25	-3.87	9.79	7.36	*5.70*
Std. Error	1.12	0.45	0.12	0.30	0.97	1.16	0.39	0.47	*2.71*
Dichotic									
Mean	3.13	-0.03	21.83	4.46	1.44	-3.14	7.21	6.95	*5.23*
Std. Error	0.66	0.83	0.10	0.41	0.72	1.29	0.93	0.34	*2.67*
Pooled Var.	6.15	4.64	0.10	1.04	8.83	12.64	7.76	1.39	*57.82*
Effect Size	0.31	0.37	-0.34	-0.70	0.27	-0.21	0.93	0.35	*0.06*
p Value	0.40	0.51	0.51	0.20	0.51	0.68	0.09	0.50	
Tone + Noise									
Inattentive									
Mean	14.36	11.72	22.14	13.54	15.53	13.97	17.62	14.16	*15.38*
Std. Error	1.80	0.94	0.69	0.33	0.46	0.64	1.35	0.88	*1.13*
Dichotic									
Mean	17.25	12.57	24.92	13.65	15.95	13.33	14.65	13.87	*15.77*
Std. Error	0.60	0.87	0.42	0.40	0.33	0.64	1.00	0.71	*1.41*
Pooled Var.	11.21	7.37	2.92	1.09	1.95	3.49	11.25	5.36	*13.36*
Effect Size	-0.86	-0.31	-1.62	-0.10	-0.30	0.34	0.89	0.13	*-0.11*
p Value	0.13	0.54	0.007	0.85	0.48	0.50	0.10	0.80	

Table 5: Average nSFOAE Magnitudes and Effect Sizes from Triplet 2 of the Inattentive- and Dichotic-Listening Conditions

Tone Alone	L01	L02	L03	L04	L05	L06	L07	L08	*M*
Inattentive									
Mean	5.31	1.89	21.92	7.21	9.05	7.02	19.62	13.72	*10.72*
Std. Error	1.18	0.77	0.04	0.77	0.60	0.56	0.29	0.70	*2.50*
Diotic									
Mean	6.59	4.10	22.15	6.85	10.81	5.45	19.25	11.74	*10.87*
Std. Error	0.41	0.79	0.13	0.61	0.84	1.07	0.47	0.69	*2.34*
Pooled Var.	6.75	4.04	0.05	3.78	4.07	3.33	0.61	3.29	*48.05*
Effect Size	-0.49	-1.10	-1.00	0.19	-0.87	0.86	0.47	1.09	*-0.02*
p Value	0.52	0.10	0.08	0.76	0.10	0.17	0.47	0.02	
Tone + Noise									
Inattentive									
Mean	14.86	11.78	22.77	11.85	16.19	17.26	21.81	14.92	*16.43*
Std. Error	2.19	0.80	0.59	0.62	0.45	0.49	0.76	0.71	*1.45*
Diotic									
Mean	18.04	13.72	24.94	12.53	14.33	14.03	18.99	12.49	*16.13*
Std. Error	1.25	0.99	0.94	0.41	0.52	1.18	1.20	1.39	*1.52*
Pooled Var.	24.44	4.77	3.94	2.34	2.15	3.11	4.12	5.15	*17.29*
Effect Size	-0.64	-0.89	-1.10	-0.45	1.27	1.83	1.39	1.07	*0.07*
p Value	0.37	0.17	0.06	0.48	0.02	0.01	0.10	0.06	

Table 6: Average nSFOAE Magnitudes and Effect Sizes from Triplet 1 of the Inattentive- and Diotic-Listening Conditions

Tone Alone	L01	L02	L03	L04	L05	L06	L07	L08	*M*
Inattentive									
Mean	3.89	0.76	21.72	3.74	2.25	-3.87	9.79	7.36	*5.70*
Std. Error	1.12	0.45	0.12	0.30	0.97	1.16	0.39	0.47	*2.71*
Diotic									
Mean	2.29	1.43	21.68	4.65	2.38	-3.09	9.47	7.13	*5.74*
Std. Error	0.50	0.54	0.14	0.54	0.90	1.15	1.39	0.74	*2.64*
Pooled Var.	6.17	1.49	0.12	0.87	9.41	10.19	3.75	1.89	*57.80*
Effect Size	0.64	-0.55	0.12	-0.97	-0.04	-0.24	0.17	0.17	*0.00*
p Value	0.34	0.38	0.82	0.15	0.93	0.69	0.80	0.74	
Tone + Noise									
Inattentive									
Mean	14.36	11.72	22.14	13.54	15.53	13.97	17.62	14.16	*15.38*
Std. Error	1.80	0.94	0.69	0.33	0.46	0.64	1.35	0.88	*1.13*
Diotic									
Mean	18.34	14.38	25.03	13.42	15.96	13.49	15.81	13.78	*16.28*
Std. Error	1.05	0.45	0.70	0.32	0.53	1.26	2.09	1.61	*1.38*
Pooled Var.	16.50	5.23	3.75	0.73	2.28	4.39	12.67	7.49	*11.88*
Effect Size	-0.98	-1.16	-1.49	0.15	-0.28	0.23	0.51	0.14	*-0.26*
p Value	0.17	0.07	0.02	0.81	0.59	0.68	0.46	0.79	

Table 7: Average nSFOAE Magnitudes and Effect Sizes from Triplet 2 of the Inattentive- and Diotic-Listening Conditions

Triplet 1	L01	L02	L03	L04	L05	L06	L07	L08	*M*	*SD*
Inattentive	47.7	66.2	X	16.0	23.5	33.8	X	14.2	**33.6**	*20.2*
Dichotic	38.9	73.9	X	26.5	10.6	30.1	X	13.1	**32.2**	*23.0*
Diotic	38.8	64.8	X	22.5	11.8	31.9	X	18.4	**31.4**	*19.0*
Triplet 2	L01	L02	L03	L04	L05	L06	L07	L08	*M*	*SD*
Inattentive	48.0	144.9	X	27.6	X	42.3	174.2	27.9	**77.5**	*64.7*
Dichotic	43.4	135.5	X	34.0	X	36.9	45.2	21.1	**52.7**	*41.5*
Diotic	41.0	118.2	X	39.1	X	36.5	53.0	24.3	**52.0**	*33.7*

Table 8: Average nSFOAE Time Constants (ms) in the Auditory-Attention Conditions

Triplet 1	L01	L02	L03	L04	L05	L06	L07	L08	*M*	*SD*
Inattentive	-8.8	-11.2	-9.2	-11.0	-10.9	-11.0	-8.0	-10.0	***-10.0***	*1.2*
Dichotic	-12.5	-12.5	-13.4	-13.4	-13.6	-13.3	-10.7	-13.2	***-12.9***	*1.0*
Diotic	-13.5	-14.8	-16.7	-15.0	-16.4	-16.9	-14.6	-18.2	***-15.8***	*1.5*

Triplet 2	L01	L02	L03	L04	L05	L06	L07	L08	*M*	*SD*
Inattentive	-10.1	-10.8	-8.8	-12.4	-11.5	-11.4	-9.4	-10.6	***-10.6***	*1.2*
Dichotic	-12.6	-12.1	-13.6	-13.2	-13.1	-13.1	-11.8	-13.2	***-12.8***	*0.6*
Diotic	-16.6	-14.8	-16.3	-16.1	-16.8	-16.0	-13.5	-14.5	***-15.6***	*1.2*

Table 9: Average Physiological Noise-Floor Magnitudes (dB) in the Auditory-Attention Conditions
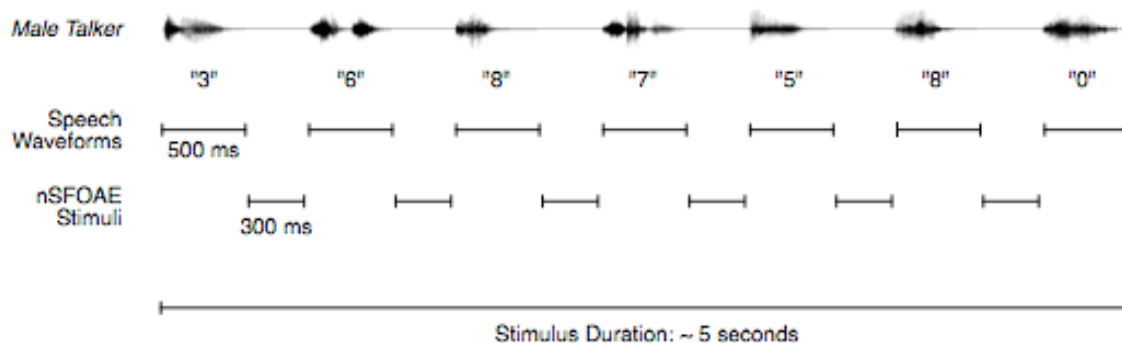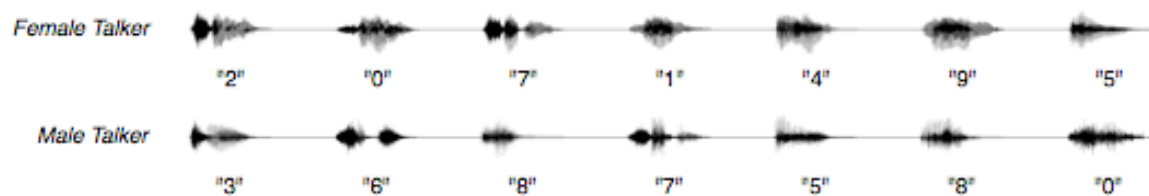
Figure 1. An example of the stimulus waveforms presented to the ears during one trial of the dichotic-listening condition. Each ear was presented with a series of seven spoken digits, one series spoken by a female talker, and the other series spoken simultaneously by a male talker. The ear receiving the female talker was selected randomly on each trial. Each digit was fitted to a 500-ms temporal window, and a 300-ms ISI separated consecutive digits. Although not shown here, the nSFOAE-eliciting stimuli were presented in the ISIs. After each trial the subject performed a matching task based upon the digits spoken by the female talker. A 2000-ms silent response interval marked the end of each trial (not shown).

Waveforms presented to one ear:
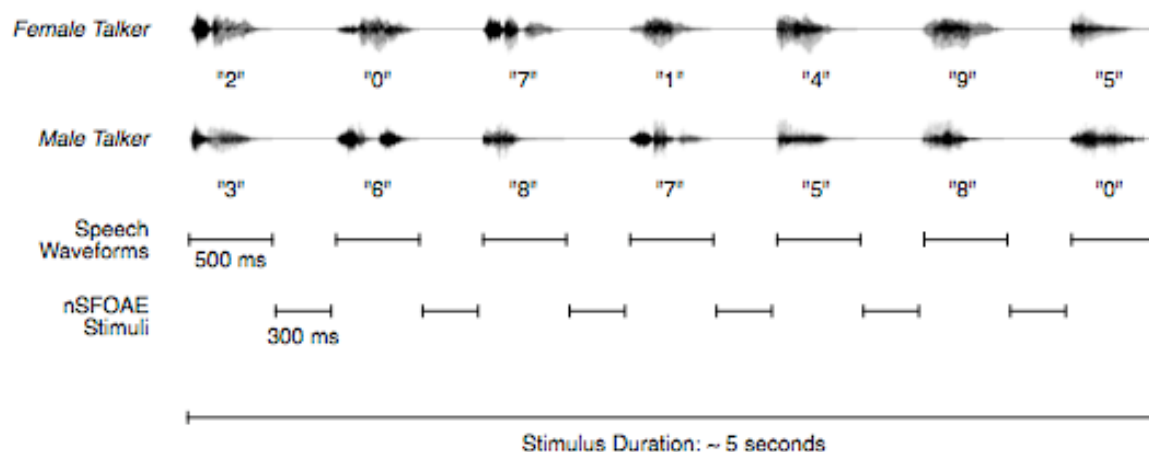

Waveforms presented to the opposite ear:

Figure 2. An example of the stimulus waveforms presented to each ear during one trial of the diotic-listening condition. Both ears were presented with two series of seven spoken digits, one from a female talker and one from a male talker. The task of the subject was the same as in the dichotic-attention condition—to attend to the female talker on each trial and to use her spoken digits to perform the subsequent forced-choice matching task.
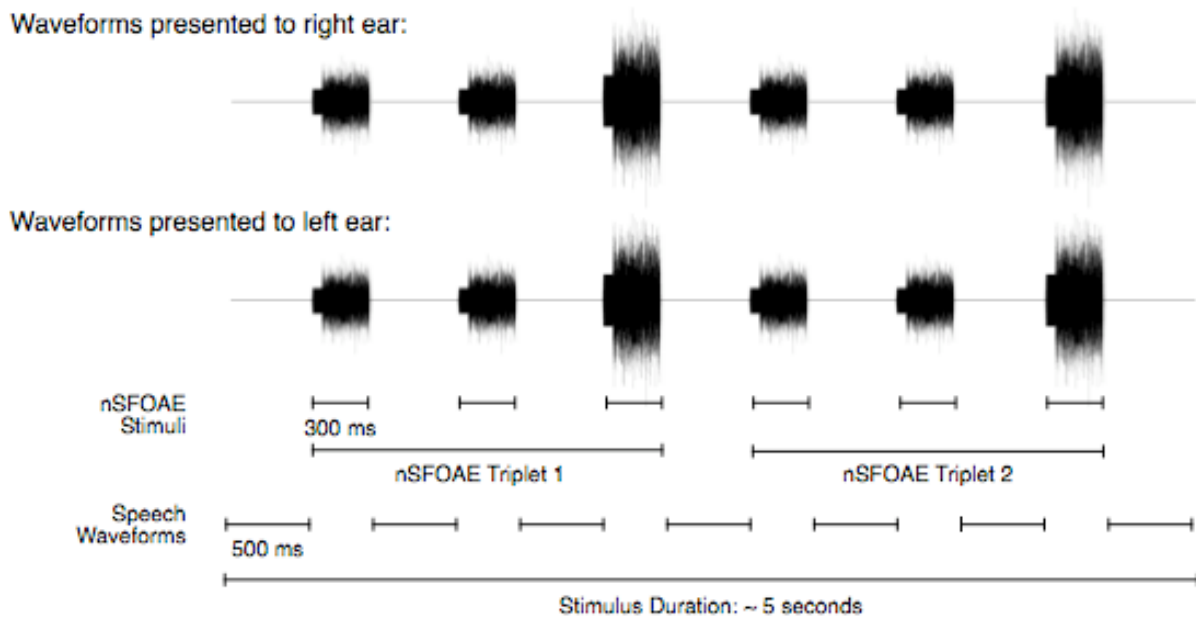
Figure 3. Schematic showing the series of stimulus waveforms presented on every trial of the nSFOAE procedure. The nSFOAE stimulus always was composed of a 300-ms tone and a 250-ms frozen wideband noise, and the onset of the tone always preceded the onset of the noise by the difference in their durations. A random ISI lasting about 500 ms separated consecutive stimuli (see text), and a 2000-ms response interval ended each trial. Although not shown here, the seven speech waveforms used for the behavioral tasks were interleaved with the nSFOAE stimuli. A cancellation procedure was performed separately on each of the two Triplets presented on each trial, yielding two estimates of the nSFOAE per trial. The nSFOAE responses from Triplet 1 and Triplet 2 always were averaged and analyzed separately. Not shown here are the 30-ms recordings in the quiet that followed the presentation of every nSFOAE stimulus that were used to estimate the physiological noise floor for every behavioral trial (see text). They were omitted here for simplicity.
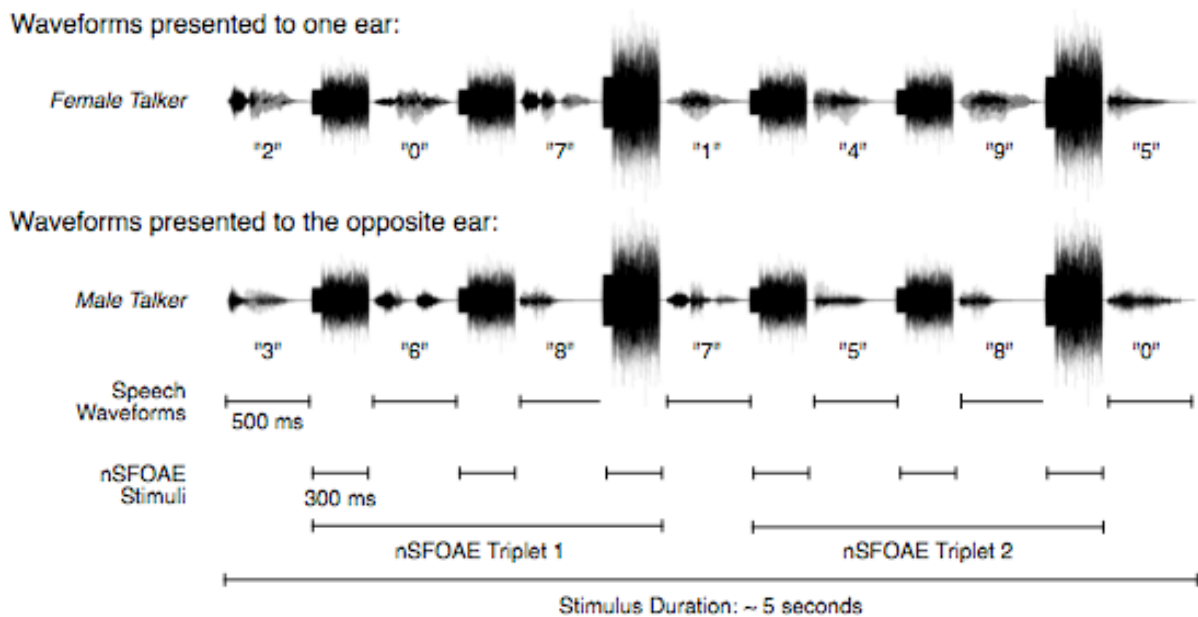
Figure 4. Schematic showing how the nSFOAE stimuli were presented at the same time that subjects' performed the dichotic-listening task. In both ears, the spoken digits were interleaved with the nSFOAE-eliciting stimuli. On each trial, nSFOAE responses were recorded from the right ear only, but the subject attended to the right or left ear with equal probability. Not shown here are the 30-ms recordings in the quiet that followed the presentation of every nSFOAE stimulus that were used to estimate the physiological noise floor for every behavioral trial (see text). They were omitted here for simplicity.
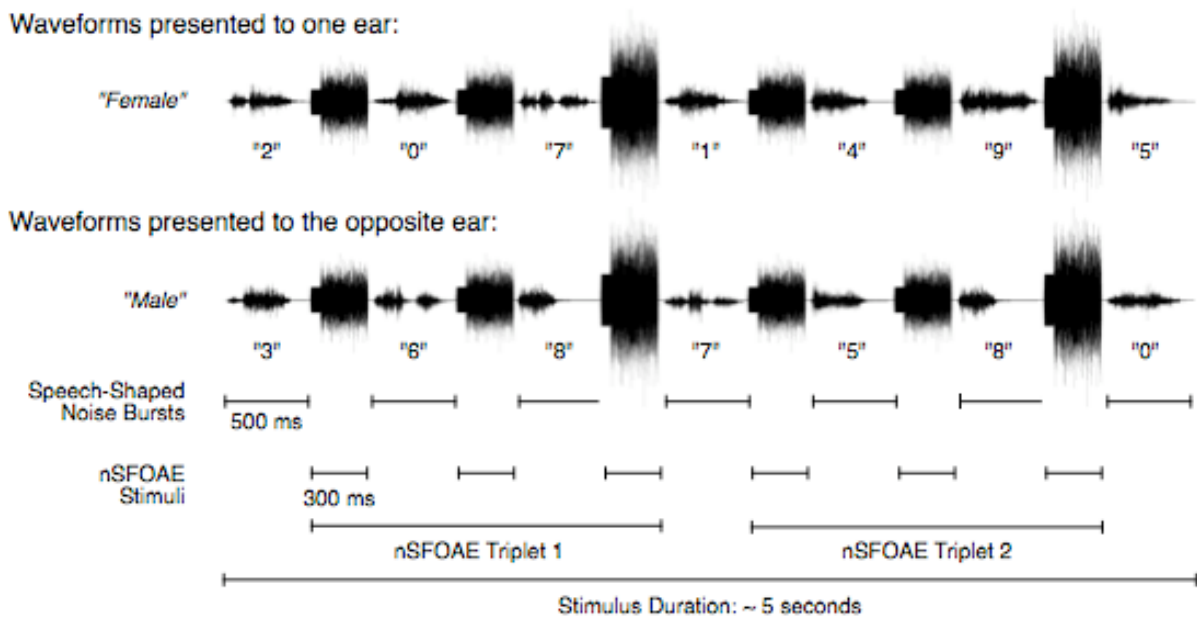
Figure 5. Schematic showing the series of stimulus waveforms presented on every trial of the inattentive-listening condition. In this condition, speech-shaped noise bursts (rather than speech waveforms) were interleaved with the same nSFOAE stimuli used in the selective-listening conditions. The speech-shaped noise bursts had the same spectra, the same temporal envelopes, and the same overall level as the speech waveforms themselves, but did not sound like speech. The task of the subject in the inattentive-listening condition simply was to press the number "4" on the response keypad at the end of final sound presented on each trial. Not shown here are the 30-ms recordings in the quiet that followed the presentation of every nSFOAE stimulus that were used to estimate the physiological noise floor for every behavioral trial (see text). They were omitted here for simplicity.
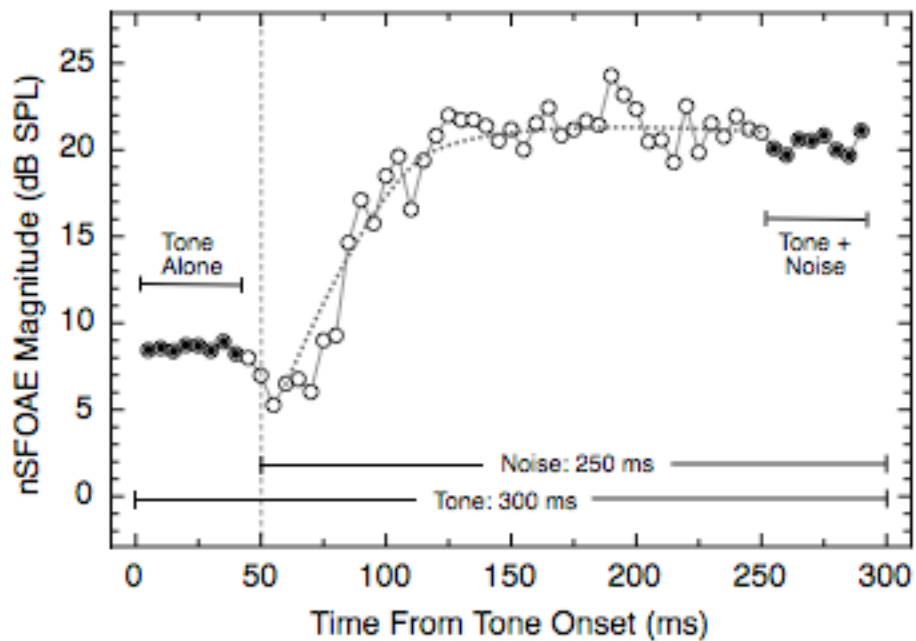
Figure 6. An example of an nSFOAE response averaged from one block of trials of one condition, and a visual description of how it was analyzed for the purposes of evaluation and comparison. For analysis, tone-alone magnitudes were averaged from 5 – 40 ms, and tone-plus-noise magnitudes were averaged from 250 – 285 ms (35-ms epochs each). A positive exponential function (dotted line) was fitted to the waveform from 60 – 250 ms, and a time constant was calculated from the fit. All nSFOAE responses, across all subjects and conditions, were analyzed in this same manner.

Figure 7. Schematic showing the steps involved in averaging and pooling the nSFOAE responses from repeated measures a single behavioral condition. The left side of the figure illustrates the individual nSFOAE waveforms from one block of trials, the middle illustrates the averaged nSFOAE responses from repeated blocks of trials, and the right side of the figure illustrates that the averaged responses were pooled to yield a single representative response for a particular condition. In this schematic, only the nSFOAEs from Triplet 1 of a condition are shown. The nSFOAEs from Triplet 2 were treated in the exact same manner.

Triplet One

Triplet Two

nSFOAE Magnitude (dB SPL)

Time From Tone Onset (ms)
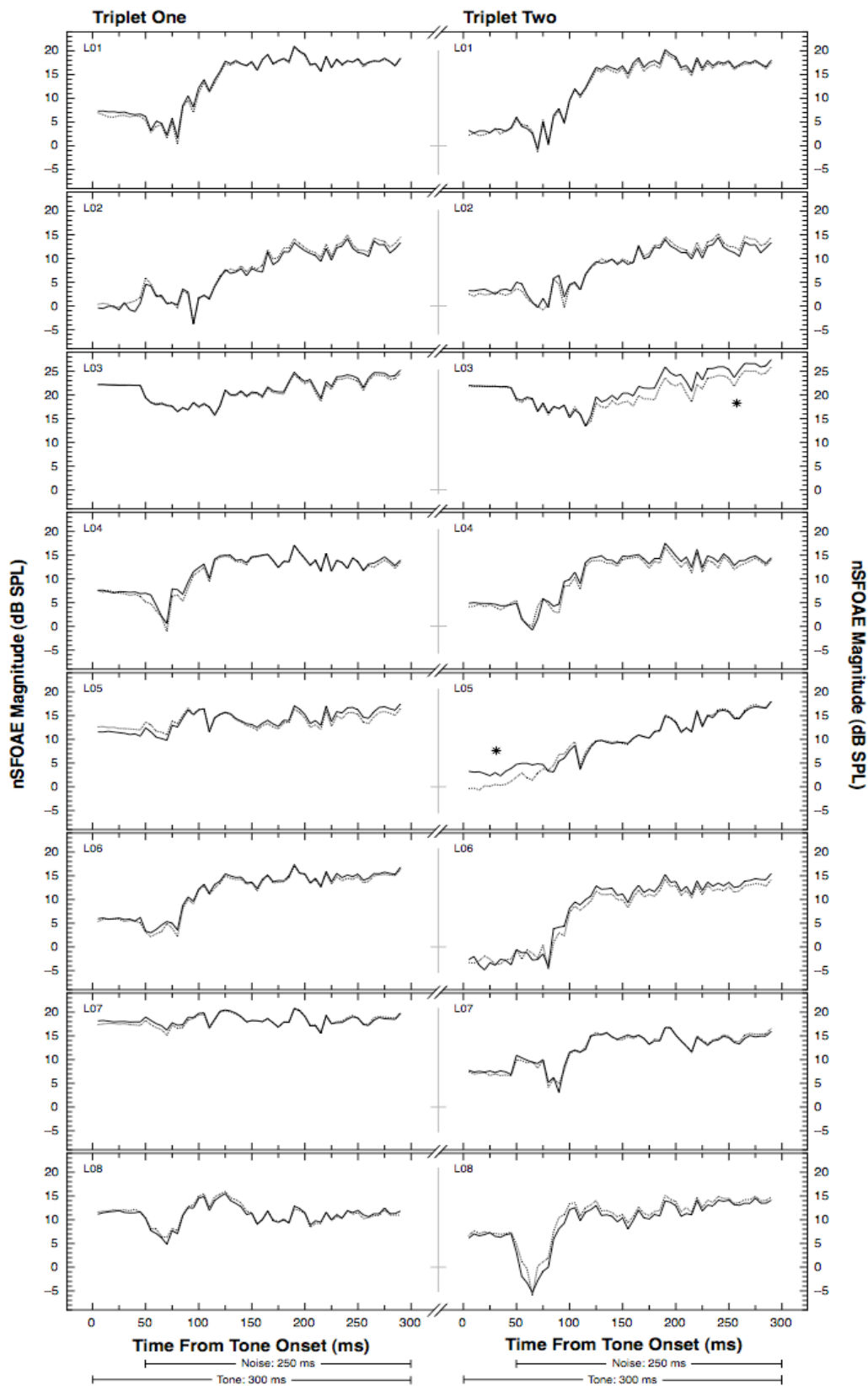
Noise: 250 ms
Tone: 300 ms

103

Figure 8: Averaged nSFOAE responses from repeated measures of the dichotic-listening condition. The data from Triplet 1 are shown on the left side of the figure, the data from Triplet 2 are shown on the right side of the figure, and each horizontal pair of panels contains the data of an individual subject. The solid line in each panel shows the averaged ipsilateral response (right ear attended), and the dashed line shows the averaged contralateral response (left ear attended). For the majority of subjects, there was no effect of differentially attending to the two ears (see text for exceptions).
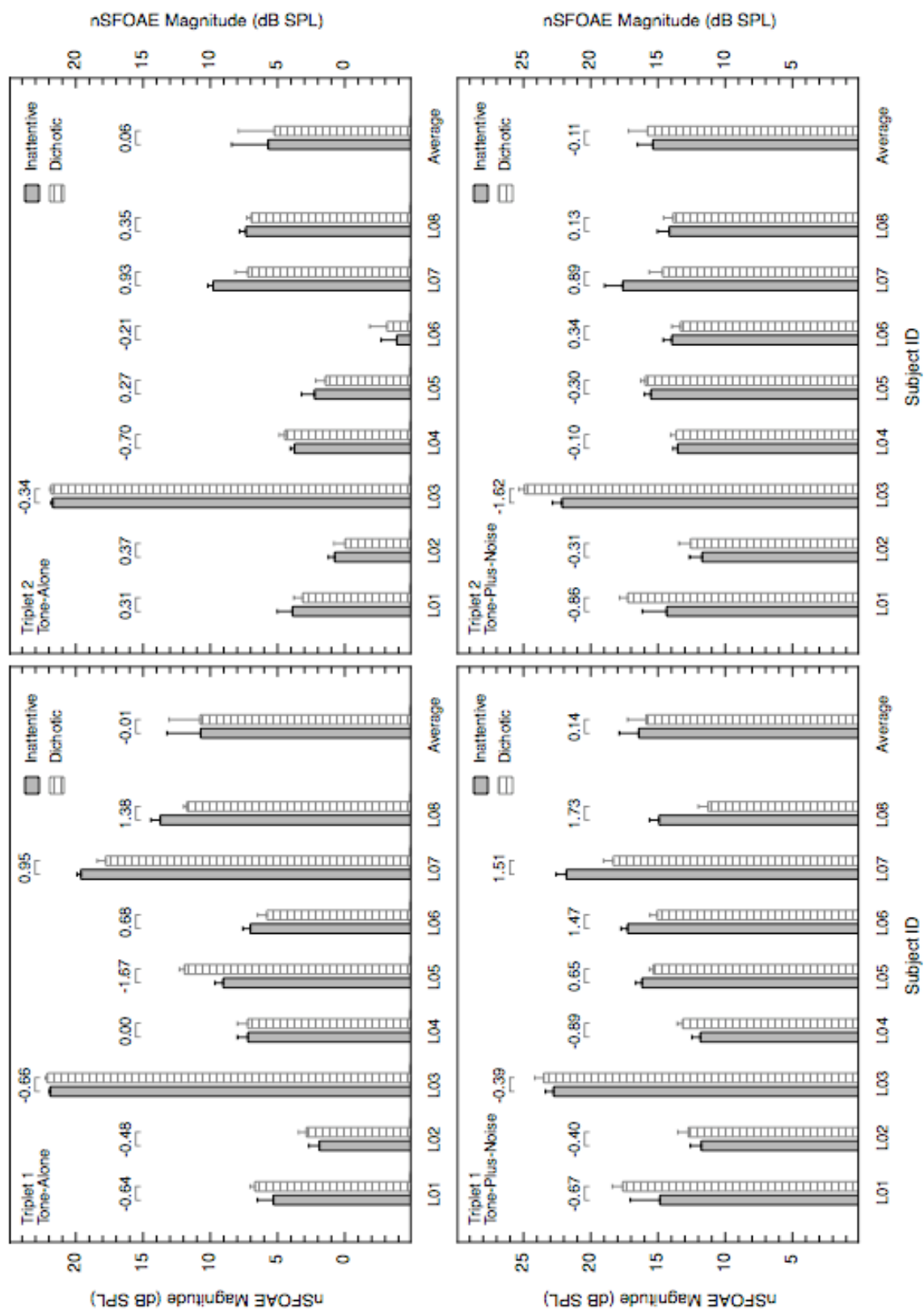
Figure 9: Average magnitudes of the nSFOAE responses from repeated measures of the inattentive-listening (grey bars) and dichotic-listening (striped bars) conditions. The tone-alone and tone-plus-noise data from Triplet 1 are shown on the left side of the figure, and the comparable data from Triplet 2 are shown on the right side of the figure. The tone-alone data are shown in the top two panels, and the tone-plus-noise data are shown in the bottom two panels. The range of ordinate values for the top and bottom panels is the same, but the scales are different by 5.0 dB. The subject ID is shown below each pair of bars, and the final pair of bars at the far right of each panel shows the average nSFOAE magnitudes calculated across all eight subjects in the two conditions. An effect size was calculated for each comparison by subtracting the average magnitude from the dichotic-listening condition from the average magnitude from the inattentive-listening condition, and dividing by a common standard deviation.
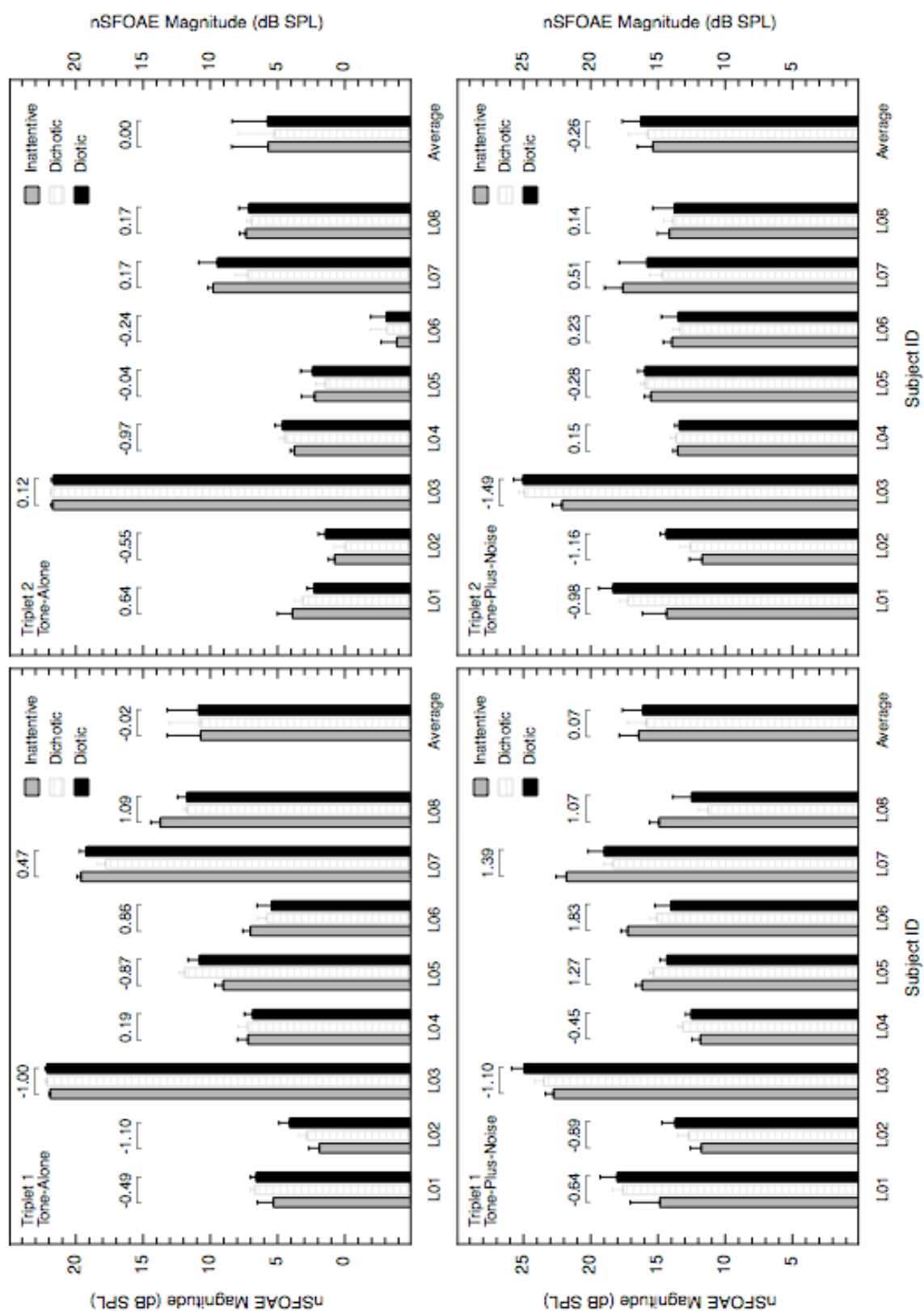
Figure 10: Average magnitudes of the nSFOAE responses from repeated measures of the diotic-listening condition, shown as black bars. The data from the inattentive-listening (grey bars) and dichotic-listening (striped bars) conditions are replotted here for comparison. As in Figure 9, the data from Triplet 1 are shown on the left side of the figure, the data from Triplet 2 are shown on the right side of the figure, the tone-alone data are shown in the top two panels, and the tone-plus-noise data are shown in the bottom two panels. The range of ordinate values for the top and bottom panels is the same, but the scales are different by 5.0 dB. The subject ID is shown below each pair of bars, and the final pair of bars at the far right of each panel shows the average nSFOAE magnitudes calculated across all eight subjects in the two conditions. An effect size was calculated for each comparison by subtracting the average magnitude from the dichotic-listening condition from the average magnitude from the inattentive-listening condition, and dividing by a common standard deviation.
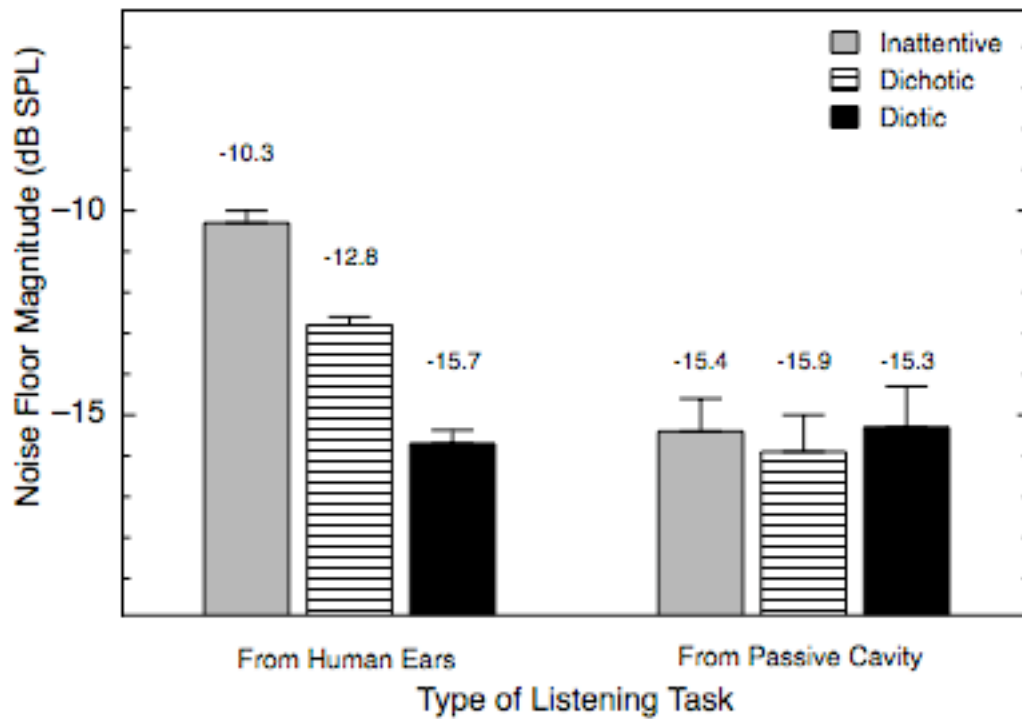
Figure 11. Average noise-floor magnitudes from the inattentive- (grey bars), dichotic- (striped bars), and diotic-listening (black bars) conditions. The data in the left half of the figure are physiological noise-floor magnitudes calculated across the eight subjects, and the data in the right half of the figure are physical noise-floor magnitudes calculated across repeated measures in a passive cavity (a syringe). Across all eight subjects, and evident in the averages shown here, noise-floor magnitudes recorded in the ear canals of our subjects were lower in the dichotic-listening condition versus the inattentive-listening condition, and lowest in the diotic-listening condition. In contrast, noise-floor magnitudes recorded in the passive cavity did not differ across conditions.

Visual-Attention Condition

Feedback

87451 87431

Response Interval

87451 87431

200 ms

3 0

2000 ms maximum

9 1

nSFOAE Stimuli

500 ms
300 ms

0 5

500 ms
300 ms

Triplet 2

8 4

500 ms
300 ms

2 7

500 ms
300 ms

Triplet 1

1 8

500 ms
300 ms

5 2

500 ms
300 ms

500 ms

Interstimulus intervals during which the nSFOAE stimuli were presented to the two ears. The digits were removed from the screen during the ISIs.

*Only the middle five digits in the series were used for the forced-choice matching task.
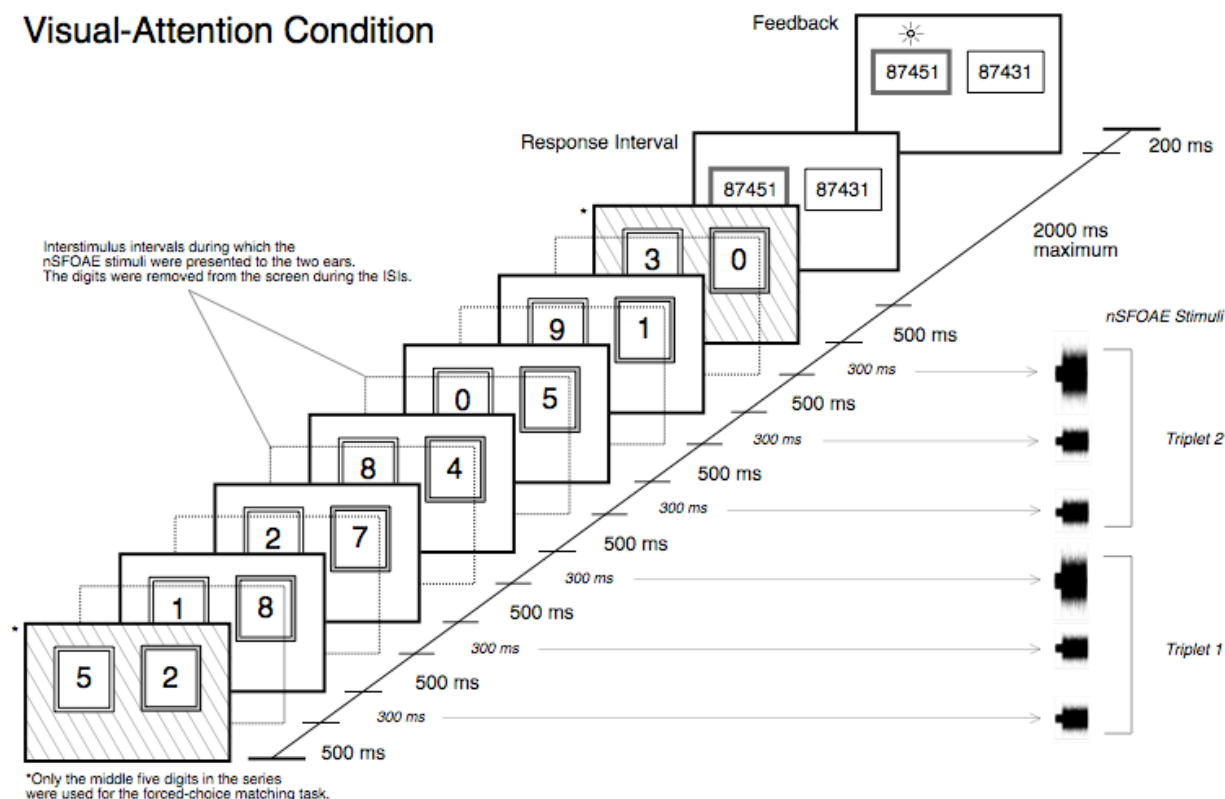
Figure 12: Schematic showing a single trial of the visual-attention condition. Two series of seven digits were presented consecutively to one of two adjacent locations on a computer screen, and the subject was instructed to attend to one set of digits—the one designated by a pink border (shown here as a grey border). On each trial, the middle five digits in the attended series were used to complete the two-alternative forced-choice matching task (see asterisks above). A 2000-ms response interval followed the last pair of digits in the series, and feedback was provided on every trial. Shown here as dashed, transparent frames, 300-ms interstimulus intervals separated consecutive pairs of digits. During these intervals the digits were removed from the computer screen, and the nSFOAE stimuli were presented to the two ears. Not shown here are the 30-ms recordings in the quiet that followed the presentation of every nSFOAE stimulus, and that were used to estimate the noise floor for every behavioral trial (see text), just as was true for the auditory-attention conditions. For the visual-attention condition with speech, the speech stimuli used for the auditory-attention conditions were presented to the ears at the same time, and for the same duration, that the digits used for the visual-attention condition were presented to the computer screen. The digits presented to the ears were unrelated to those presented visually, and could not be used to complete the behavioral task. For the visual-attention condition without speech, only the nSFOAE stimuli were presented to the ears.

References

Avan, P., and Bonfils, P. (1992). "Analysis of possible interactions of an attentional task with cochlear micromechanics," Hear. Res. 57, 269 – 275.

Bacon, S. P., Fay, R. R., and Popper, A. N. (Eds.). (2004). Compression: From cochlea to cochlear implants. New York: Springer.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (1985). "Evoked mechanical responses of isolated cochlear outer hair-cells," Science 227, 194 – 196.

Collet, L., and Duclaux, R. (1986). "Auditory brainstem evoked responses and attention," Acta Otolaryngol. 101, 439 – 441.

Connolly, J. F., Karen, A., McGillivary, N., and Scott, D. W. (1989). "Human brainstem auditory evoked potentials fail to provide evidence of efferent modulation of auditory input during attentional tasks," Psychophysiol. 26, 292 – 303.

Davis, H. (1983). "An active process in cochlear mechanics," Hear. Res. 9, 79 – 90.

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.

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. (1993a). "Transiently evoked otoacoustic emission amplitudes change with changes of directed attention," Physiol. Behav. 53, 679 – 682.

Froehlich, P., Collet, L., Valatx, J. L., and Morgon, A. (1993b). "Sleep and active cochlear micromechanical properties in human subjects," Hear. Res. 66, 1 – 7.

Galambos, R. (1956). "Suppression of auditory-nerve activity by stimulation of efferent fibers to the cochlea," J. Neurophysiol. 19, 424 – 437.

Geisler, W. S., and Cormack, L. (2011). "Models of overt attention." In S. P. Liversedge, I. D. Gilchrist, and S. Everling (Eds.), *Oxford Handbook of Eye Movements*. New York: Oxford University Press.

Giard, M-H., Collet, L., Bouchet, P., and Pernier, J. (1994). "Auditory selective attention in the human cochlea," Brain Res. 633, 353 – 356.

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., 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.

Guinan, J. J., Jr., and Gifford, M. L. (1988a). "Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. I. Rate-level functions," Hear. Res. 33, 97 – 114.

Guinan, J. J., Jr., and Gifford, M. L. (1988a). "Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. II. Spontaneous rate," Hear. Res. 33, 115 – 128.

Guinan, J. J., Jr., and Gifford, M. L. (1988a). "Effects of electrical stimulation of efferent olivocochlear neurons on cat auditory-nerve fibers. III. Tuning curves and thresholds at CF," Hear. Res. 37, 29 – 46.

Guinan, J. J., Jr., Warr, J. B., and Norris, B. E. (1983). "Differential olivocochlear projections from lateral versus medial zones of the superior olivary complex," Comp. Neurol. 221, 358 – 370.

Hall, J. W. III (1992). "Handbook of auditory evoked responses," Boston: Allyn and Bacon.

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.

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.

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.

Kawase, T., and Liberman, M. C. (1993). "Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones," J. Neurophysiol. 70, 2519 – 2532.

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

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467 – 477.

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

Maison, S., Micheyl, C., Chays, A., and Collet, L. (1997). "Medial olivocochlear system stabilizes active cochlear micromechanical properties in humans," Hear. Res. 133, 89 – 98.

Maison, S., Micheyl, C., and Collet, L. (2001). "Influence of focused auditory attention on cochlear activity in humans," Psychophys. 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. (1993). "Comparative influence of repeated measurements and of attention on evoked otoacoustic emissions," Acta Otolaryngol. 113, 471 – 477.

Meric, C., and Collet, L. (1994a). "Attention and otoacoustic emissions: A review," Neurosci Biobehav. Rev. 18, 215 – 222.

Meric, C., and Collet, L. (1994b). "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.

Picton, T. W., and Hillyard, S. A. (1971). "Human auditory evoked potentials. II. Effects of attention," Electroenceph. and Clinical Neurophys. 36, 191 – 199.

Picton, T. W., Hillyard, S. A., Galambos, R., and Schiff, M. (1971). "Human auditory attention: A central or peripheral process," Science. 173, 351 – 353.

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

Robertson, D., Anderson, C-J., and Cole, K. S. (1987). "Segregation of efferent projections to different turns of the guinea pig cochlea," Hear. Res. 25, 69 – 76.

Smith, R.L. (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," J. Acoust. Soc. Am. 65, 166-178.

Smith, R.L., and Zwislocki, J.J. (1975). "Short-term adaptation and incremental responses in single auditory-nerve fibers," Biol. Cybern. 17, 169-182.

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.

Warr, W. B., and Guinan Jr., J. J. (1979). "Efferent innervation of the organ of Corti: Two separate systems," Brain Res. 173, 152 – 155.