

Published in final edited form as:

Hear Res. 2014 June ; 312: 160–167. doi:10.1016/j.heares.2014.03.013.

Selective attention reduces physiological noise in the external ear canals of humans. II: Visual attention

Kyle P. Walsh^{a,b,c}, Edward G. Pasanen^b, and Dennis McFadden^b

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

Abstract

Human subjects performed in several behavioral conditions requiring, or not requiring, selective attention to visual stimuli. Specifically, the attentional task was to recognize strings of digits that had been presented visually. A nonlinear version of the stimulus-frequency otoacoustic emission (SFOAE), called the nSFOAE, was collected during the visual presentation of the digits. The segment of the physiological response discussed here occurred during brief silent periods immediately following the SFOAE-evoking stimuli. For all subjects tested, the physiological-noise magnitudes were substantially weaker (less noisy) during the tasks requiring the most visual attention. Effect sizes for the differences were >2.0 . Our interpretation is that cortico-olivo influences adjusted the magnitude of efferent activation during the SFOAE-evoking stimulation depending upon the attention task in effect, and then that magnitude of efferent activation persisted throughout the silent period where it also modulated the physiological noise present. Because the results were highly similar to those obtained when the behavioral conditions involved auditory attention, similar mechanisms appear to operate both across modalities and within modalities. Supplementary measurements revealed that the efferent activation was spectrally global, as it was for auditory attention.

Keywords

selective attention; otoacoustic emission; cross-modality effects; cochlear noise

1. Introduction

The data reported here reveal differences in the cochlea when human subjects are selectively attending, or not attending, to visual stimuli. The strong implication is that the medial olivocochlear (MOC) efferent system produces differing amounts of inhibition on relevant cochlear elements depending upon the difficulty of the visual-attention task being

© 2014 Elsevier B.V. All rights reserved.

^aCorresponding Author: Kyle P. Walsh, kpwalsh@umn.edu.

^cCurrent Address: Department of Psychology and Center for Cognitive Sciences, 75 East River Road, University of Minnesota, Minneapolis, Minnesota 55455

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

performed. The results are highly similar to those reported in a companion paper when the behavioral task required auditory, not visual, attention. In that previously described work (Walsh et al. 2014), subjects listened to strings of seven digits and then were required to indicate which of two visually presented digit arrays consisted of the middle five digits in the auditory presentation. As a complication, one male and one female talker spoke different stimulus digits simultaneously, and the subjects were required to attend only to the female voice. Interleaved with the successive digit pairs were other auditory stimuli which permitted the collection of our cochlear response. Thus, the cochlear response was obtained as simultaneously as possible with the presentation of the behaviorally relevant digit stimuli. One primary result from our previous report was that a measure of the noise level in the ear canal was weaker when the subjects were attending to the auditory digits than when they were not. In the current report, the behavioral task required attention to strings of *visually* presented digits but otherwise the procedures were highly similar to those in the auditory task. The results also were highly similar; a measure of physiological noise differed depending upon the degree of selective attention required for the behavioral task.

When planning these auditory- and visual-attention studies, we adopted a procedure used previously to good effect when studying the cochlear underpinnings of an auditory phenomenon known as overshoot (Walsh et al., 2010a, 2010b). This procedure is a variant on the double-evoked technique developed by Keefe (1998); its strength is that it provides a sensitive measure of the nonlinear behavior of the cochlea *during* the presentation of the stimuli used to evoke that response. To distinguish this perstimulatory measure from other measures of the stimulus-frequency otoacoustic emission (SFOAE), we have called our nonlinear measure the nSFOAE. In fact, the *perstimulatory* measurements made during the nSFOAE-evoking stimuli did covary substantially with the level of selective attention, and those results will be reported elsewhere, but for these initial companion papers on auditory and visual attention, the focus is on the cochlear responses measured during a brief silent period at the end of each nSFOAE-evoking stimulus.

2. Methods

2.1. General

The data described here were collected as part of the same study, with the same subjects, in which the auditory-attention data were collected (Walsh et al., 2014). In all cases, the auditory-attention conditions were tested several weeks prior to the visual-attention conditions. The Institutional Review Board at The University of Texas at Austin approved the procedures described here. All subjects provided their informed consent prior to any testing, and they were paid for their participation. First we describe the behavioral task for visual attention, then the procedures for measuring the nSFOAE, and finally how the behavioral and physiological measures were interleaved.

Subjects—Two males (both aged 22) and five females (aged 20 – 25) participated in the visual-attention phase of the larger study, which required two 2-hr test sessions. All of these seven subjects previously served in a highly similar auditory-attention study (Walsh et al., 2014). All subjects had normal hearing [≤ 15 dB Hearing Level (HL)] at octave frequencies between 250 and 8000 Hz, and normal middle-ear and tympanic reflexes, all determined

using an audiometric screening device (Auto Tympanometer 38, GSI/VIASYS, Inc., Madison, WI). No subject had a spontaneous otoacoustic emission (SOAE) stronger than -15.0 dB SPL within 600 Hz of the frequency of the 4 kHz probe tone.

2.2. Behavioral measures

Each subject was tested individually while seated in a reclining chair inside a double-walled, sound-attenuated room. Sounds were delivered directly to the two external ear canals using two insert earphone systems, which are described in detail in Walsh et al. (2014). A computer screen was used to provide the visual stimuli plus task instructions and trial-by-trial feedback. This screen was attached to an articulating mounting arm that was positioned by the subject to a comfortable viewing distance. Subjects provided behavioral responses by pressing designated keys on a numerical keypad.

2.2.1. Selective visual-attention conditions—The visual-attention conditions were designed to be similar to the auditory-attention conditions reported previously (Walsh et al., 2014). They differed primarily in their requirement that the subjects had to base their behavioral responses on visual rather than auditory stimuli. There were four stimulus conditions, all measured in separate blocks of trials. A block of trials continued until there were a prescribed number of artifact-free nSFOAE responses (see Walsh et al., 2014 for details), so a block could last anywhere from about 4 to about 6 minutes. Across sessions, each subject completed each of the experimental conditions to be described a minimum of four times.

In the two visual-attention conditions, subjects saw two simultaneous sequences of digits at adjacent locations on the computer screen and then were required to indicate which of two static visual arrays of digits matched what had been presented initially. For one of these attention conditions, the only *auditory* stimuli presented was a series of six sounds (described below) used to evoke the nSFOAE response. For the other attention condition, seven pairs of spoken digits were interleaved with the nSFOAE-evoking sounds and thus were simultaneous with the visually presented digits. One talker for each pair of sounds was female, the other was male; one voice was in the left ear and the other in the right at random (dichotic stimulation); and the digit strings they spoke were unrelated to each other and to the strings being shown visually.

More specifically, in the visual-attention conditions, two simultaneous sequences of digits were presented visually on each trial. Each digit was in black typeface on a white rectangular background surrounded by a larger, colored border. The border around one rectangle was pink (nominally “female”), the other border was blue (nominally “male”), the subject always was instructed to attend to the sequence of digits that was presented within the rectangle with the pink border, and, on each trial, the probability that the “pink” digits were located on the left side of the screen was equal to their being on the right side of the screen. There were seven digits in each sequence, selected randomly with replacement on every trial, and selected independently for the pink and blue series. When sequences of auditory digits also were presented, they coincided temporally with the sequences of visual digits. Each visual digit appeared on the screen for 500 ms, and each successive digit except

the last in the series was followed by a 330-ms interval that contained an nSFOAE-evoking stimulus. Thus, there were a total of six nSFOAE-evoking stimuli per trial (two “triplets,” see below).

In the two *inattention* conditions, digit sequences were not presented to the computer screen, only the rectangular, colored borders within which the digits were displayed during the two visual-attention conditions. In one inattention condition, only the nSFOAE-evoking sounds were presented to the ears, and in the other inattention condition, additional sounds (speech-derived sounds; see below) were interleaved with the nSFOAE stimuli. In both conditions, the sole requirement of the subject was to press a response key after the last sound of each trial was presented. The only difference between the inattention conditions in the visual- and auditory-attention studies was the presence of the colored borders used for the visual display in the current study.

At the end of each trial in the conditions *requiring* selective visual attention, two static series of five digits each were presented side-by-side for 2000 ms. One series was the same as the middle five digits presented sequentially within the pink border, the other series had one digit different from what was presented, and the subject had 2000 ms to press one of two response keys to indicate which series was correct. The correct series was located to the left on half of the trials. At the end of the response interval, the correct series was indicated visually (trial-by-trial feedback). At the end of each trial in the conditions *not* requiring selective attention, the subject was required only to press a single response key immediately after the final sound was presented to the ears. No response alternatives were shown. The timing and structure of these trials was as similar as possible to those used for the auditory-attention study (Walsh et al., 2014). Figure 1 illustrates the timing sequence of the visual digits and nSFOAE-evoking stimuli (described below) for a single trial in the visual-attention conditions.

Speech-shaped noise (SSN) as distractor sounds: As noted, in one of the two visual-attention conditions, spoken digits were presented simultaneously with the visual digits as distraction sounds. Distraction sounds also were used in one of the two inattention conditions, but these were not actual spoken digits. The worry was that subjects would not be able to ignore spoken digits; intentionally or not, they *would* be attending to something even in the inattention condition. Accordingly, distraction sounds were constructed by taking the Fast Fourier Transform (FFT) of each of the 20 speech waveforms used in the auditory-attention conditions (Walsh et al., 2014), discarding the phase information, and taking the inverse FFTs of those spectra, thereby creating 20 samples of noise having the same long-term frequency and amplitude spectra as the original digits. Those waveforms then were modulated with the temporal envelopes of the corresponding, original digits. Those temporal envelopes were extracted using a Hilbert transform and then lowpass filtered at 500 Hz to limit the moment-to-moment fluctuations. Thus, these speech-shaped noise (SSN) stimuli had the same long-term spectral characteristics, envelopes, and durations as the spoken digits, and they did sound different from each other, but they did not sound like speech.

The series of SSN stimuli presented to one ear on one trial all came originally from the female talker, and the sounds in the other ear all came from the male talker, but which ear received the female series was randomized trial-by-trial. Different series of SSN stimuli, selected randomly, were presented on different trials. The SSN stimuli were presented simultaneously with the visual digits in one of the two inattention conditions. While these manipulations were important as controls, they were not especially salient for the subjects, who were not required to attend preferentially to any of the auditory stimuli. It is intuitive that the visual-attention condition also having the spoken digits present should be the most demanding behaviorally because it required ignoring both the blue-bordered string of visual digits and the conflicting auditory input; in the end, however, those auditory stimuli proved not to be especially distracting.

2.3. Physiological measures

The stimuli used to evoke the nSFOAE responses were presented on the same trials used to collect the behavioral responses, just as in the auditory study. The nSFOAE-evoking stimuli and their presentation are described in detail in the companion paper (Walsh et al., 2014). Briefly, for 50 ms a 4.0-kHz tone (60 dB SPL) was presented alone; then for 250 ms a weak noise (bandwidth 0.1 – 6.0 kHz; spectrum level about 25 dB SPL) was added to the tone; then 30-ms of silence followed. The same double-evoked procedure described by Walsh et al. was used to extract the physiological responses, and to cancel the much stronger stimulus waveforms in the process. As with the auditory-attention study, only those physiological responses meeting explicit criteria were accepted for addition with the other responses of that block. Subjects received trial-by-trial feedback as to whether their physiological responses were rejected for being too noisy, and an additional trial was added to the block of trials when a response was rejected; this encouraged our subjects to sit quietly during data collection.

As a check on the software, hardware, and the procedures, the various stimulus conditions were played into a syringe having a volume of approximately 0.5 cm³, using the same earphone and microphone system and testing programs as used in the right ear for the humans. As expected, the double-evoked algorithm yielded no evidence of an nSFOAE response. Linearity prevailed. These exercises also produced estimates of the acoustic and electronic noise floor of our system; it was approximately -13.2 dB SPL at the frequency of our nSFOAE-evoking stimulus (4.0 kHz). In addition, they provided control measures to identify any possible artifacts that might create spurious differences in the measurements across conditions.

2.3.1. Physiological-noise measure—The analyses of the physiological-noise data were performed exactly as they were for the auditory-attention experiment (Walsh et al., 2013), and only will be described briefly here. At the end of each block of trials, there existed four averaged nSFOAE responses: one each for the behaviorally correct trials for triplet 1, the incorrect trials for triplet 1, and the correct and incorrect trials for triplet 2. (Because the averaged responses from incorrect trials were based on so few trials, their analyses are not presented here.) Again, for the sake of efficiency, we sometimes refer to these averaged residuals from the linear prediction as averaged difference-waveforms and

sometimes as just the nSFOAE. The difference-waveforms were 330 ms in duration. An example is shown at the top of Fig. 2. The first 50 ms represents the perstimulatory nSFOAE to the tone alone, the next 250 ms was the perstimulatory nSFOAE to the tone-plus-noise, and the final 30 ms was obtained during the silent periods that followed immediately after each nSFOAE-evoking stimulus.

For most analyses, our interest was not on the entire wideband cochlear response, but only on the response in a relatively narrow frequency region. Accordingly, for our analyses we used a bandpass filter centered at a selected frequency, with a bandwidth equal to 10% of the center frequency (a 6th-order, elliptical digital filter; 400 Hz wide and centered at 4.0 kHz, for example). Specifically, the initial 10-ms segment of the 330-ms difference-waveform was filtered, and the rms voltage was calculated, transformed into decibels SPL, and stored. Then the 10-ms analysis window was advanced 1 ms, and the rms voltage of the filtered waveform was calculated again, transformed, and stored. This process continued in 1-ms steps up through the final 10-ms window in the silent period, beginning at 319 ms from the start of the difference-waveform. As noted, elsewhere we will report the perstimulatory results; here the focus is on the levels of the averaged filtered responses obtained during the silent period. As shown at the bottom of Fig. 2, the levels of the difference-waveforms often declined gradually over the course of the first few milliseconds of silence (revealing persistence of the cochlear response to the tone-plus-noise stimulus) before reaching an asymptotic level. Accordingly, for the analyses below, the averages obtained from the first 10 analysis windows of the silent period are separated from those from the final 10 windows. (As shown in Walsh et al., 2014, the time courses and magnitudes of these gradual declines differed across subjects and across frequency, and the likelihood is that they arise from decaying SFOAEs.)

2.4. Data analysis

Each block of at least 30 trials yielded two usable wideband averaged difference-waveforms (using only trials with correct behavioral responses), one for triplet 1 and one for triplet 2. Those difference-waveforms were analyzed as shown in Fig. 2. In order to summarize the data, the results for individual subjects were pooled across blocks by averaging the sequences of levels calculated, 10-ms-window-by-10-ms-window. For each subject, for each experimental condition, at least four averaged, filtered difference-waveforms contributed to those sequences of means (corresponding to about 80 – 120 individual trials). In the end, the data from the silent period were highly similar for triplets 1 and 2 (as was true in the auditory study, Walsh et al. 2014), so for simplicity, only the former are reported here. Finally, the decibel values for the ten successive windows beginning 10 ms after the onset of the silent period were averaged to characterize the asymptotic level of the physiological noise reached (see below).

Physiological-noise magnitudes from the selective-attention and inattention conditions were compared using analyses of variance (ANOVA), and the differences between experimental conditions were summarized here using effect size, or *d* (see Cohen, 1992). Effect size is the difference between the means of two distributions of data divided by an estimate of the pooled standard deviation across the two distributions. By convention, an effect size

between 0.20 – 0.50 is considered to be small, an effect size between 0.50 – 0.80 is considered to be medium, and an effect size greater than 0.80 is considered to be large (Cohen, 1992).

3. Results

3.1. Visual selective attention

Subjects performed well on the behavioral digit-recognition task in both of the selective visual-attention conditions, indicating that they were attending reliably to the correct stimuli. This is an important outcome to consider when interpreting the differences between physiological-noise levels in the attention and inattention conditions. In the two visual-attention conditions (with and without the spoken distraction stimuli), on average, subjects performed at 91.1% correct (range = 78.7% – 99.2%). There was no statistical difference in behavioral performance between the two tasks. Unfortunately, the number of trials with behavioral errors was too few to support believable separate analyses, so only the data from correct trials will be presented here.

The physiological-noise levels measured during the silent periods were substantially higher in the inattention conditions than in the visual selective-attention conditions. As indicated at the bottom of Fig. 2, the magnitudes of the physiological responses declined over the first few milliseconds of the silent period and then reached an asymptote; the values for the inattention conditions were larger than those for the attention conditions throughout this time course. For simplicity, and for consistency with the auditory-attention study (Walsh et al., 2014), the levels in the final ten 10-ms analysis windows, beginning 10 ms into the silent period, were averaged within subjects and then across subjects for each experimental condition. Because there was no statistical difference between physiological-noise levels for the two triplets, only the data for triplet 1 are shown here. The mean data from the two control conditions (without and with the SSN distraction stimuli) are shown in the left half of Table 1, and the data from the two corresponding visual-attention conditions are shown in the right-most columns. Individual-subject data are shown in successive rows, and the means and standard deviations calculated across subjects are shown below those data. For consistency with the auditory study, the values shown in the table and figure here have been adjusted by 2.46 dB to correct for the rise-time of the digital filter at 4.0 kHz (see Walsh et al., 2014).

As can be seen from Table 1, the physiological-noise levels measured under attentional load were weaker than in the inattention conditions, and this was true for every subject. Both findings are highly similar to the results reported for auditory attention using parallel procedures (Walsh et al., 2014). The differences of about 2 – 3 dB between conditions were statistically significant: a one-way ANOVA revealed a significant main effect of condition on the average noise magnitudes [$F(3, 24) = 19.0, p < 0.0001$]. The effect sizes for the differences between attention and inattention *without* the auditory distraction sounds and attention and inattention *with* the auditory distraction sounds were 3.2 and 2.5, respectively. In contrast, when the acoustic stimuli from these four conditions were played to, and recorded from, a syringe (a non-human, passive cavity), using exactly the same equipment,

software, and procedures as was used with the human subjects, the physical noise measures were roughly equal in all conditions, about -13.2 dB at 4.0 kHz.

Also evident from Table 1 is that the speech or SSN stimuli affected noise levels only slightly. Comparing column 2 with column 3, and column 4 with column 5 in Table 1 reveals that the levels of the physiological noise typically were slightly smaller when the speech or SSN stimuli were presented along with the visual digits. Perhaps actual speech would have been more distracting than the SSN stimuli in the inattention conditions, but that would have created the potential problem of the subjects attending to the actual speech when the goal was for them to be attending relatively little – simply listening for the last sound presented on each trial, or attending to the visual cues on the computer screen, in order to determine the start of the response interval.

One interpretation of these results is that efferent control of cochlear responses was modulated during the behavioral tasks by those cortical centers involved in processing attended stimuli, whether visual (present report) or auditory (as in Walsh et al., 2014). Similar orderings and time courses were seen across conditions in the auditory-attention study (see Figures 5 – 8 in Walsh et al., 2014). Also similar to the auditory data, when the difference-waveforms obtained from corresponding moments of the silent period in triplets 1 and 2 of the visual-attention and inattention conditions were compared, the correlations were in the range of 0.8 – 0.9 for the first 10-ms analysis window during the silent period before declining to values of about 0.2 – 0.4 beginning about 7 ms into the silent interval. These declines in correlation obtained were seen for both the visual-attention and inattention conditions. Together, these facts suggest that there were stimulus-related aftereffects in the ear canal (i.e., echoes) that persisted into the silent period, but only for a few milliseconds before dying away. The high correlations between the difference-waveforms reveal that these aftereffects were highly similar on every presentation of the nSFOAE-evoking stimuli. This would be true if the after effects were decaying SFOAEs to the final moments of the frozen samples of tone and wideband noise. These declines were not observed in recordings obtained from a passive cavity, confirming that they were not attributable to the decay of the stimuli themselves. Once the declines were complete, the physiological noises asymptoted to the levels shown in Table 1.

3.2. Across-frequency effects

Physiological-noise measures from the silent period also were made at a number of other frequencies; the frequencies were the same as those selected for comparable analyses in Walsh et al. (2014). The procedure was as follows: For each subject and each condition, the averaged noise waveform for triplet 1 was bandpass filtered symmetrically around each selected frequency, using a filter whose bandwidth was 10% of that center frequency. Physiological-noise magnitudes at each frequency were calculated for each subject by averaging the levels in the final ten analysis windows during the silent period (from 10 – 19 ms), just as was done for the measures in Table 1. Those values then were averaged across the seven subjects. Finally, those latter means were corrected to cancel out the effects of differential rise times in the filters (see Walsh et al., 2014).

The results for two conditions are shown as solid symbols with solid lines in Fig. 3. The conditions shown are the visual-attention and inattention conditions that *did* include speech or SSN distractor stimuli. Also shown in Fig. 3 (open symbols with dotted lines) are similar data from the same subjects extracted from the responses in the auditory-attention (dichotic speech) and inattention conditions (Walsh et al., 2014). As can be seen, noise magnitudes in all conditions were largest at the lowest and highest frequencies tested, and smaller at intermediate frequencies. More importantly, for every frequency examined, and for both the visual- and auditory-attention tasks, the magnitude of the physiological noise was stronger in the inattention condition than in the selective-attention condition. The similarity in the results across modalities is remarkable. These across-subject means are fully representative of the results for individual subjects. Moreover, parallel outcomes were obtained for the corresponding conditions in which the speech or SSN distraction sounds were *not* present. The strong implication is that the mechanisms involved in producing the attentional effects were activated about equally strongly for visual and auditory tasks, and also that those mechanisms were operating globally across the spectrum. The dashed and solid lines in Fig. 3 are described below.

The physiological-noise magnitudes of the individual subjects were compared across the four visual-attention conditions using a one-way univariate ANOVA, with experimental condition and center frequency (of the bandpass filter) as the two factors. Significant main effects were revealed for both condition [$F(3, 288) = 63.8, p < 0.0001$] and center frequency [$F(11, 288) = 156.6, p < 0.0001$], but not for the interaction of the two factors [$F(33, 288) = 0.2, p = 1.0$].

In Fig. 3, note that the differences between attention conditions were essentially the same at the frequency of the relatively strong 4.0-kHz tone (that was required for measuring the perstimulatory nSFOAE) as at neighboring frequencies. Also note that the two highest frequencies shown in Fig. 3 (6.5 and 7.0 kHz) were outside the passband of the frozen sample of wideband noise that was presented simultaneously with the 4.0-kHz probe tone, yet the attentional effects were the same.

For comparison purposes, also shown in Fig. 3 (dashed line) are estimates of the electronic noise floor of our recording system, as measured in a passive cavity (a 0.5-cc syringe). These were collected using the same test programs as were used for the subjects and also extracted with the same analysis programs. The values shown are the averages of data collected with the software appropriate for each of the four visual test conditions. Just as was true for the auditory study (Walsh et al., 2014), and just as should be true here, there were no evident differences in the estimates of the electronic noise floors obtained using those different test programs. Had the increases across frequency in either the noise floor or the magnitudes of the physiological noise been wholly attributable to the proportional (10%) bandwidths used for the analysis filters, the slopes of the functions would have been approximately the same as the slope of the straight line shown at the bottom right in the Fig. (3 dB per octave). The additional steepness of the functions, particularly at the highest frequencies tested, can be explained by the overall operating characteristic of our recording system, which exhibits a rise of about 7 – 8 dB over the range from 3.5 to 7.0 kHz.

Finally, physiological-noise levels also were measured at the frequencies of each subject's two strongest SOAEs. (Five of the seven subjects had SOAEs in the right ear that were strong enough to be detected.) Just as was observed across the frequency spectrum (Fig. 3), SOAE magnitudes during the silent period also typically were larger during inattention than during selective visual attention, although the differences between conditions typically were somewhat smaller than at the other tested frequencies. On average, SOAE magnitudes were about 1.0 – 1.5 dB larger during the inattention conditions compared to the visual-attention conditions. At SOAE frequencies, within-subject correlations between pairs of averaged waveforms from triplet 1 and 2 of a condition were high (about 0.8 – 0.9) throughout the entire 30-ms silent period, indicating that the SOAEs were synchronized by the preceding tone-plus-noise nSFOAE-evoking stimulus.

4. Summary

1. A tone-plus-noise stimulus was used to activate the MOC efferent system and to obtain a nonlinear measure of cochlear response (the nSFOAE) during both the perstimulatory period and a 30-ms silent period following the tone-plus-noise stimulus. In some conditions, the subjects needed to attend to visual stimuli in order to perform a behavioral task; in other conditions, they had no reason to attend.
2. Physiological-noise magnitudes measured from the ears of human subjects were significantly weaker during selective visual attention than during inattention for both triplet 1 and triplet 2, and for all seven subjects tested. The magnitudes of the differences were about 2.5 – 3.0 dB on average, and this was true whether spoken digits or speech-shaped noise was presented simultaneously to the ears during the presentation of the visual digits.
3. For some subjects, noise magnitudes declined during the first few milliseconds of the silent period before reaching an asymptote, indicating that there were echo-like responses (decaying SFOAEs) to the preceding tone-plus-noise stimuli that persisted into the silent period. Throughout the entire time course, however, magnitudes were weaker during attention than inattention.
4. The same pattern of results was observed across the frequency spectrum, indicating that the underlying physiological mechanisms operated globally in the cochlea for the stimuli used here. At each selected frequency, physiological-noise levels were lower during visual attention than during inattention.
5. Cross correlations between pairs of averaged waveforms from triplet 1 and 2 of the same condition were high at the start of the silent period (following stimulus offset), and then declined to about zero during the last ten analysis windows (from which our average noise measure was obtained), indicating that using the nSFOAE cancellation procedure during the end of the silent period was tantamount to summing three uncorrelated samples of noise (see below).
6. When our (wideband) noise measures were filtered around the frequencies of subjects' SOAEs, high correlations (0.8 – 0.9) between averaged waveforms were observed *throughout* the silent period, and with no shift in relative time alignment

between the waveforms, indicating that subjects' SOAEs were being synchronized by the preceding tone-plus-noise stimulus used to evoke the perstimulatory nSFOAE responses.

7. All of the above outcomes are consistent with the results from a companion study that used selective auditory- rather than visual-attention tasks (Walsh et al., 2014), indicating that the observed effects of attention operate *across* perceptual modalities, in addition to operating within the auditory modality. This is in accord with the conclusion reached first by Hernández-Peón et al. (1956).

5. Discussion

The results presented here show that selective visual attention reduces the level of physiological noise measured from the cochleae of human subjects. In a companion paper, the same outcome was observed during selective auditory attention in a comparable behavioral task (Walsh et al., 2014). In both studies, physiological-noise levels measured during the inattention conditions were highly similar, and this might have been expected because the inattention conditions differed only in minor aspects of the visual display. Unexpected across studies, however, was the similarity of the noise levels during the various selective-attention conditions, and, consequently, the size of the effect of attention. The sensory modality used to perform the digital-recall task apparently had no differential effect on the MOC efferent activity. How could this be? At the cognitive level, the task for the subject was essentially the same in both experiments: sequences of seven digits were presented at the same temporal rate, the attentional load and the difficulty of the task were the same, and the behavioral motor response required at the end of every trial was the same. The only difference was the mode of presentation of the digit sequences. If the efferent activity originated from a multimodal stage of perceptual processing, that activity might well have been independent of the mode of presentation. Note also that the task for the subject was predominantly cognitive in nature. The presentation of the test digits, whether visual or auditory, did not involve any challenge in visual or auditory processing. In other words, there was no additional demand on the visual or auditory system to acquire the digits that eventually would be tested.

5.1. Interpreting cochlear noise

The nSFOAE as described here is, by definition, a measure of nonlinearity, and that nonlinearity almost surely originates from the cochlea because measurements in a passive cavity reveal no detectable activity above the residual noise floor of the measurement system. Such a measure of nonlinearity as a waveform is a challenge to interpret, but when applied to the final milliseconds of the silent interval, we believe that the interpretation becomes simpler. Once the aftereffects of the nSFOAE-evoking stimuli had died out, the double-evoked subtraction during the silent period was tantamount to summing two independent samples of noise and subtracting a third independent sample, perhaps of the same approximate magnitude. In other words, we believe that we were summing *three* independent samples of noise (further discussed in Walsh et al., 2014). This outcome is somewhat easier to interpret because in this case the measure is no longer an estimate of deviation from linearity; it is simply a measure of noise magnitude. The reduction in the

observed noise level under attention could be interpreted in this context quite simply as increased efferent activity reducing displacement on the basilar membrane (Cooper and Guinan, 2006).

In the companion paper (Walsh et al., 2014, Section 5.1), we have stated our belief that there was cochlear-generated noise in the ear canal during the final moments of the silent periods and that that noise was being modulated differentially (by the persisting MOC activation from the preceding tone-plus-noise stimulus) depending upon the attentional condition. We believe that the same processes are at play here. The present experiment differs only in the visual (not aural) presentation of the test stimuli for the behavioral task. We have outlined in the companion paper our interpretations of the results and included speculations concerning the origins of the cochlear noise that we were able to measure. We also explain in the companion paper why noise from body or head movements is unlikely to be the source of our physiological-noise measures from the ear canal. While we cannot say for sure from these data *what* the acquired benefit might be for the listener under attentional demand, one possibility is that attention improves the signal-to-noise ratio (SNR) for target sounds presented to the ears simply by decreasing the level of the background noise that competes with the focus of the listener's attention.

5.1.1. Dependence of observed noise level upon number of samples averaged

—The averaged waveforms that were analyzed to obtain noise levels were nominally each based on about the same number of trials. For example, in this study, the average number of trials for the inattention condition with speech-shaped noise was 32.0 (SD = 4.9), and for the visual-attention condition with speech it was 32.1 (SD = 4.0). The duration of a block of trials also was similar in the inattention and attention conditions. For example, the average block duration for the inattention condition with speech-shaped noise was 268.4 s (SD = 75.1), and for the visual-attention condition with speech it was 265.5 s (SD = 66.8). These data indicate that extraneous subject noise sufficient to reject individual responses was similar across conditions. Percent-correct performance was consistently high across test conditions, but the percent correct for inattention blocks was typically the highest, resulting in larger numbers of trials averaged for those blocks. Note that this would have worked to *reduce* the observed noise levels for inattention conditions. Hence, the effect of attention on noise level might have been slightly greater if the number of correct trials had been fixed across conditions.

5.2. Comparison with previous reports

Several previous studies have used otoacoustic emissions (OAEs) as a measure of cochlear physiology during tasks that required visual attention, or relative inattention. Some of these studies used only a visual-attention task (Puel et al., 1988; Froehlich et al., 1990; Meric and Collet, 1994), and others used both auditory- and visual-attention tasks during different blocks of trials (Froehlich et al., 1993; de Boer and Thornton, 2007). Of those studies, to our knowledge, only Froehlich et al. (1993) and de Boer and Thornton (2007) measured the levels of background noise in their OAE recordings, and only the latter study found a main effect of experimental condition on noise level. De Boer and Thornton (2007) measured click-evoked OAEs (CEOAEs) during two different visual-attention conditions, requiring

differing amounts of cognitive resources. In the “passive” visual condition, subjects watched a silent movie, and in the “active” visual condition they judged whether simple additive sums presented on a computer screen were correct. In the control, “inattention” condition, subjects pressed a button each time a noise was presented to the contralateral ear.

The results showed that the average noise level in subjects' ear canals was significantly higher during the inattention task than during the passive visual-attention task, but was no different from the average noise level during the active visual-attention task. The differences across conditions all were less than 1.0 dB SPL. With what is perhaps a complementary result, de Boer and Thornton (2007) also observed that the number of rejected trials due to excessive noise in their OAE recordings was greater during the inattention task than during the passive visual task, but not during the active visual task, which had a higher rejection rate than the inattention condition. In other words, the same directionality of the differences across conditions was observed both in this supplemental measure and in the noise measure. None of the differences in rejection rates, however, were statistically significant.

When interpreting the data from de Boer and Thornton (2007), it is important to mention that the inattention and active visual tasks each required a motor response from their subjects (a key-press), but the passive visual-attention task did not. Interestingly, there were more rejected trials and higher noise levels when a motor response was required during OAE measurement. Other previous studies showed differences in OAE *magnitudes* during attention and inattention tasks, but did not control for equivalent motor behavior in all tasks (Puel et al., 1988; Avan and Bonfils, 1992; Meric and Collet, 1992, 1994; Froehlich et al., 1993; Ferber-Viart, 1995). In the present study, differences between attention and inattention conditions cannot be attributable to different motor responses in the different behavioral conditions because a key-press was required at the end of every trial in every condition.

5.3. Final Comment

Students of the auditory system long have realized that cochlear hearing loss carries a double-whammy. When outer hair cells (OHCs) are lost, not only is there a loss of sensitivity of up to about 40 dB, there also is the loss of whatever the MOC efferent system contributes to everyday hearing. The results reported here help delineate the performance characteristics of the efferent system under conditions of selective attention. Unfortunately, however, the known details are still too few to conclude if the mechanisms documented here actually help with perception or, if so, how. The mechanism described here operated equally on the attended and the non-attended ears, operated globally across the spectrum, showed persistence (sluggishness), and operated essentially identically whether the required selective attention was visual or auditory. As noted previously (Walsh et al., 2014), these are not characteristics obviously suited to rapid changes in attention across ears or across frequency regions. Although the present data do show us that MOC efferent activity is *correlated* with selective attention, they do not reveal much about the why of the second whammy.

Acknowledgments

This study was done as part of the requirements for a doctoral degree from The University of Texas at Austin, by author KPW, who now is located at the University of Minnesota. This work was supported by a research grant awarded to author DM by the National Institute on Deafness and other Communication Disorders (NIDCD; RO1 DC000153). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCD or the National Institutes of Health.

References

- Avan P, Bonfils P. Analysis of possible interactions of an attentional task with cochlear micromechanics. *Hear Res.* 1992; 57:269–275. [PubMed: 1733917]
- Cohen J. A power primer. *Psych Bull.* 1992; 112:155–159.
- Cooper NP, Guinan JJ Jr. Efferent-mediated control of basilar membrane motion. *J Physiol.* 2006; 576:49–54. [PubMed: 16901947]
- de Boer J, Thornton ARD. Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. *Hear Res.* 2007; 233:117–123. [PubMed: 17910996]
- Ferber-Viart C, Duclaux R, Collet L, Guyonnard F. Influence of auditory stimulation and visual attention on otoacoustic emissions. *Physiol Behav.* 1995; 57:1075–1079. [PubMed: 7652027]
- Froehlich P, Collet L, Chanal JM, Morgon A. Variability of the influence of a visual task on the active micromechanical properties of the cochlea. *Brain Res.* 1990; 508:286–288. [PubMed: 2306619]
- Froehlich P, Collet L, Morgon A. Transiently evoked otoacoustic emission amplitudes change with changes of directed attention. *Physiol Behav.* 1993; 53:679–682. [PubMed: 8511172]
- Hernández-Peón R, Scherrer H, Jouvet M. Modification of electric activity in cochlear nucleus during “attention” in unanesthetized cats. *Science.* 1956; 123:331–332. [PubMed: 13298689]
- Keefe DH. Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence. *J Acoust Soc Am.* 1998; 103:3489–3498.
- Meric C, Collet L. Visual attention and evoked otoacoustic emissions: a slight but real effect. *Int J Psychophysiol.* 1992; 12:233–235. [PubMed: 1639669]
- Meric C, Collet L. Differential effects of visual attention on spontaneous and evoked otoacoustic emissions. *Int J Psychophysiol.* 1994; 17:281–289. [PubMed: 7806471]
- Puel JL, Bonfils P, Pujol R. Selective attention modifies the active micromechanical properties of the cochlea. *Brain Res.* 1988; 447:380–383. [PubMed: 3390709]
- Walsh KP, Pasanen EG, McFadden D. Properties of a nonlinear version of the stimulus-frequency otoacoustic emission. *J Acoust Soc Am.* 2010a; 127:955–969. [PubMed: 20136218]
- Walsh KP, Pasanen EG, McFadden D. Overshoot measured physiologically and psychophysically in the same human ears. *Hear Res.* 2010b; 268:22–37. [PubMed: 20430072]
- Walsh KP, Pasanen EG, McFadden D. Selective attention reduces physiological noise in the external ear canals of humans. I: Auditory attention Submitted to *Hear Res* July 2013. 2014

Abbreviations

| | |
|---------------|---|
| ANOVA | analysis of variance |
| BP | bandpass |
| CEOAE | click evoked otoacoustic emission |
| <i>d</i> | Cohen's <i>d</i> , effect size |
| HL | hearing level |
| MOC | medial olivocochlear |
| nSFOAE | nonlinear stimulus-frequency otoacoustic emission |

| | |
|---------------|---|
| OAE | otoacoustic emission |
| OHC | outer hair cell |
| SFOAE | stimulus-frequency otoacoustic emission |
| SOAE | spontaneous otoacoustic emission |
| SSN | speech-shaped noise |
| dB SPL | decibels sound-pressure level |
| Hz | hertz |
| kHz | kilohertz |
| ms | millisecond |

Highlights

Physiological noise was weaker during selective visual attention than inattention.

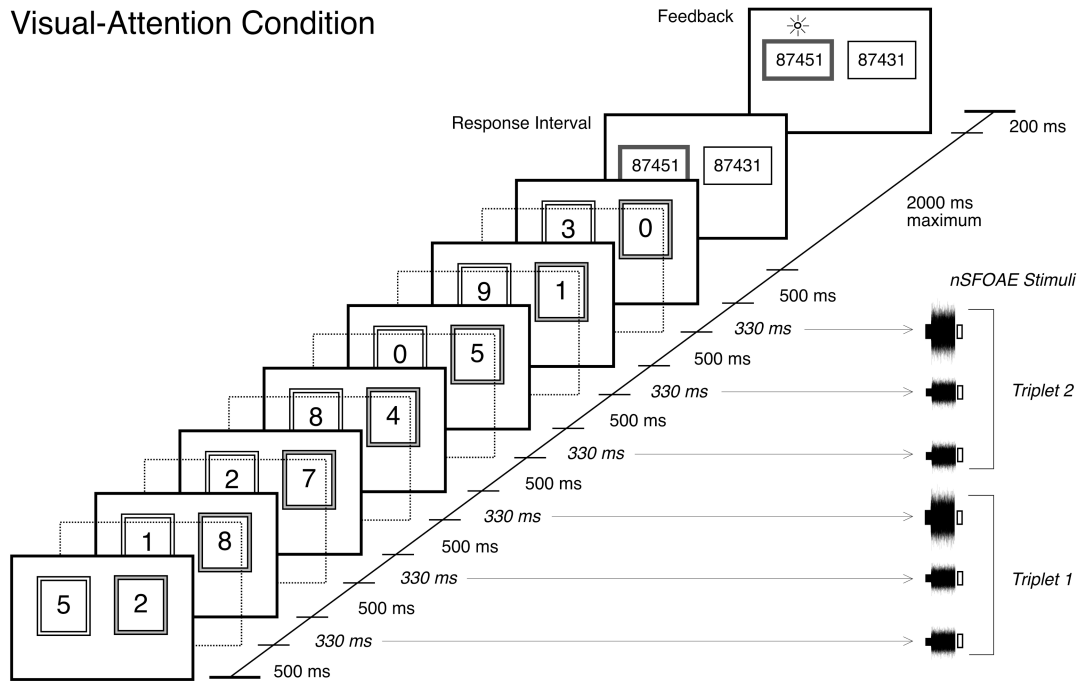
The effects of attention were evident across the frequency spectrum.

The effects of attention also were evident at subjects' SOAE frequencies.

Physiological-noise levels were similar for attention to visual and auditory stimuli.

The effects of attention on physiological noise operate across perceptual modalities.

Visual-Attention Condition

**Fig. 1.**

Schematic showing stimulus presentation during a single trial of the visual-attention condition without spoken distractors. Two series of seven digits were presented simultaneously to two adjacent locations on a computer screen, and the subject was instructed to attend to that set of digits surrounded by the pink border (shown here as a grey border). Following the last pair of digits on each trial, there was a 2000-ms response interval, during which two adjacent, static arrays of five digits each were shown, and the subject pressed one of two response keys to indicate which array was the middle five digits in the attended series. After the response, visual feedback was provided about the correct choice. The dotted, transparent frames mark the 330-ms interstimulus intervals that separated consecutive pairs of digits. During these intervals the digits were removed from the computer screen, and the nSFOAE-evoking stimuli (tone-alone followed by tone-plus-noise) were presented to the two ears diotically. No acoustic stimulation was presented during the final 30 ms of each 330-ms nSFOAE-evoking interval; those silent periods are shown on the right side as open rectangles. It is the physiological noise responses during those silent periods that are the focus of this report. For the visual-attention conditions with spoken distractors, those speech stimuli were presented separately to the ears at the same time (dichotic presentation), and for the same duration, as the digits used for the visual-attention condition were presented on the computer screen. The digits presented to the ears were unrelated to those presented visually, and could not be used to perform the behavioral task.

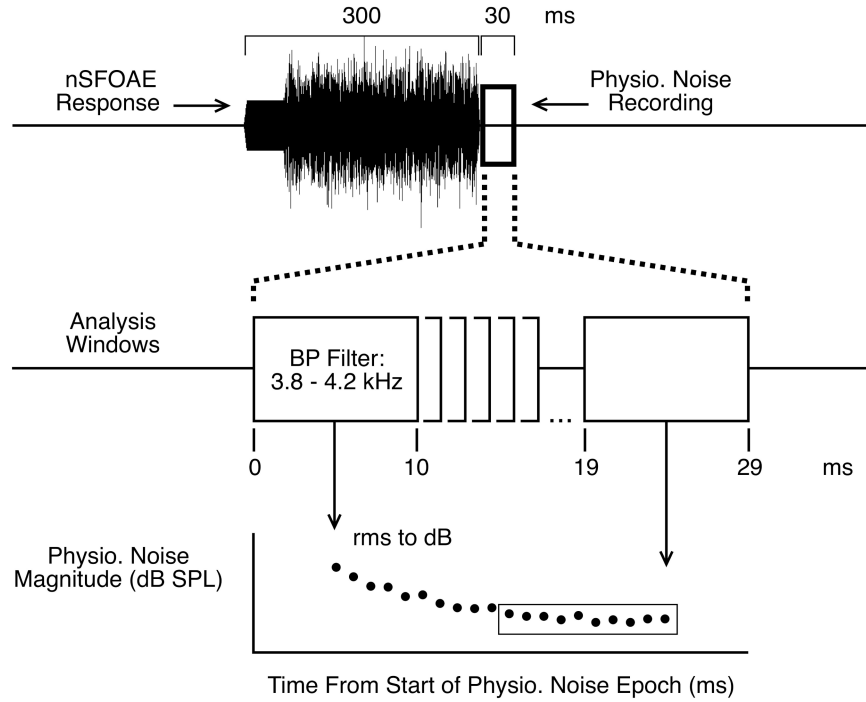


Fig. 2. Schematic showing how the physiological-noise measure was calculated. The average nonlinear physiological-noise recording from each triplet was analyzed by passing a 10-ms rectangular window through the raw waveform in 1-ms steps. At each step the physiological-noise waveform was bandpass (BP) filtered, and the rms voltage was converted to decibels SPL. The graph at the bottom of the figure shows the magnitude of the difference-waveform for the analysis windows extending from 300 to 319 ms after tone onset (the silent period). The figure emphasizes the results from the final ten windows of the 30-ms silent interval (enclosed by the rectangle), but the same analysis was performed for the entire 330-ms difference-waveform (the results for the prestimulatory segment will be reported elsewhere). (Figure reproduced from Walsh et al., 2014.)

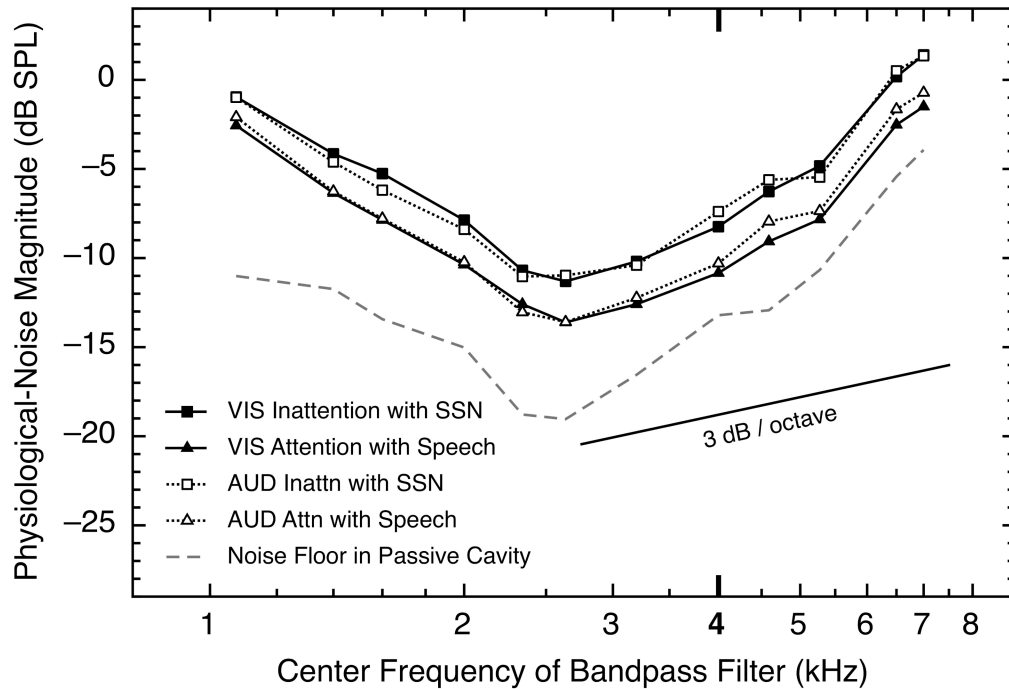


Fig. 3.

Physiological-noise magnitude as a function of frequency for four experimental conditions, all involving distractor sounds. For each data point, each of the seven subjects contributed one mean across the final ten analysis windows during the silent period (10-ms windows with onsets from 10 to 19 ms after stimulus offset), and those values were averaged across four blocks for each subject and then across seven subjects. Two of the means were obtained from blocks of trials from the auditory-attention study (Walsh et al., 2014), and two were obtained in the present study. The same seven subjects contributed to both data sets. The data from the inattention conditions are shown as squares, and the data from the selective-attention conditions are shown as triangles. At each frequency, the physiological-noise waveforms were bandpass filtered using a bandwidth that was 10% of that particular frequency. All final means were corrected for the different rise times of the 10% filters at the different center frequencies. Also shown are estimates of the electronic noise floor of our system across frequency (dashed line). A line with slope of 3 dB per octave is shown; had the increases in magnitude of the physiological noise been attributable solely to bandwidth increases in the analysis filter, the data functions would have had that slope.

Table 1

Physiological noise levels (dB SPL) in the visual-attention and inattention conditions (triplet 1 only), from trials having correct behavioral responses.

| Subject | Inattention conditions | | Visual-attention conditions | |
|---------------------------------------|------------------------|----------|-----------------------------|-------------|
| | w/o speech or SSN | with SSN | w/o speech or SSN | with speech |
| L01 | -8.8 | -8.0 | -10.3 | -10.8 |
| L02 | -7.0 | -8.4 | -10.5 | -10.7 |
| L03 | -7.1 | -8.6 | -11.2 | -10.8 |
| L05 | -8.2 | -7.6 | -9.9 | -10.9 |
| L06 | -7.7 | -9.2 | -10.4 | -12.4 |
| L07 | -7.3 | -6.3 | -8.7 | -8.8 |
| L08 | -6.0 | -9.4 | -11.3 | -11.2 |
| Mean | -7.4 | -8.2 | -10.3 | -10.8 |
| Std. Dev. | 0.9 | 1.0 | 0.9 | 1.1 |
| Effect Size (Inattention – Attention) | | | 3.2 | 2.5 |

Each entry is a window-by-window mean across at least four 30-trial blocks, corresponding to about 80 – 120 trials (correct only) averaged for each physiological response.

Only the data from triplet 1 are shown; the data from triplet 2 were highly similar.

In the inattention condition, “SSN” denotes the speech-shaped noise (SSN) distraction sounds; in the attention condition, “speech” denotes spoken digits.