Non-auditory Influences on the Auditory Periphery

by

Kurtis George Gruters

Department of Psychology and Neuroscience
Duke University

Date:_____

Approved:

_____
Jennifer Groh, Supervisor

_____
Marty Woldorff

_____
Henry Yin

_____
Tobias Overath

_____
Jeffrey Beck

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School
of Duke University

2016

ABSTRACT

Non-auditory Influences on the Auditory Periphery

by

Kurtis George Gruters

Department of Psychology and Neuroscience
Duke University

Date:_____
Approved:

_____
Jennifer Groh, Supervisor

_____
Marty Woldorff

_____
Henry Yin

_____
Tobias Overath

_____
Jeffrey Beck

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School of
Duke University

2016

Copyright by
Kurtis George Gruters
2016

Abstract

Once thought to be predominantly the domain of cortex, multisensory integration has now been found at numerous sub-cortical locations in the auditory pathway. Prominent ascending and descending connection within the pathway suggest that the system may utilize non-auditory activity to help filter incoming sounds as they first enter the ear. Active mechanisms in the periphery, particularly the outer hair cells (OHCs) of the cochlea and middle ear muscles (MEMs), are capable of modulating the sensitivity of other peripheral mechanisms involved in the transduction of sound into the system. Through indirect mechanical coupling of the OHCs and MEMs to the eardrum, motion of these mechanisms can be recorded as acoustic signals in the ear canal. Here, we utilize this recording technique to describe three different experiments that demonstrate novel multisensory interactions occurring at the level of the eardrum. 1) In the first experiment, measurements in humans and monkeys performing a saccadic eye movement task to visual targets indicate that the eardrum oscillates in conjunction with eye movements. The amplitude and phase of the eardrum movement, which we dub the Oscillatory Saccadic Eardrum Associated Response or OSEAR, depended on the direction and horizontal amplitude of the saccade and occurred in the absence of any externally delivered sounds. 2) For the second experiment, we use an audiovisual cueing task to demonstrate a dynamic change to pressure levels in the ear when a sound is

expected versus when one is not. Specifically, we observe a drop in frequency power and variability from 0.1 to 4kHz around the time when the sound is expected to occur in contract to a slight increase in power at both lower and higher frequencies. 3) For the third experiment, we show that seeing a speaker say a syllable that is incongruent with the accompanying audio can alter the response patterns of the auditory periphery, particularly during the most relevant moments in the speech stream. These visually influenced changes may contribute to the altered percept of the speech sound. Collectively, we presume that these findings represent the combined effect of OHCs and MEMs acting in tandem in response to various non-auditory signals in order to manipulate the receptive properties of the auditory system. These influences may have a profound, and previously unrecognized, impact on how the auditory system processes sounds from initial sensory transduction all the way to perception and behavior. Moreover, we demonstrate that the entire auditory system is, fundamentally, a multisensory system.

Contents

List of Figures

Acknowledgements

I am eternally grateful for the patience and valuable feedback of my advisor Jennifer Groh; committee members, Marty Woldorff, Henry Yin, Tobias Overath, and Jeffrey Beck; as well as research advisors Christopher Shera, David Smith, and Nell Cant. This work would not have been possible without the guidance and advice of this fantastic group of scientists.

1. Introduction

Organisms gather information about their environment from a variety of sensory systems, but how these systems interact with each other is poorly understood. Multisensory integration is sufficiently common in cortical regions that the cortex has been described as a fundamentally multisensory processor (for review, see Ghazanfar & Schroeder, 2006). In contrast, relatively little attention has been given to the subcortical systems that provide sensory information to the cortex. Much of the sensory information passed along to the cortex from subcortical areas is already multisensory in nature, and is influenced by behavioral state and relevance of stimuli.

The auditory system in particular is heavily influenced by multisensory inputs; how, when, and where multisensory processes first begin to influence the system is vital for understanding how the system works, and the answers to these questions are still unknown. The extensive ascending and descending connectivity of the system (figure 1) indicate that multisensory processes may begin at the very periphery of the system. The implication of this is vital for understanding auditory processes: while it is typically assumed that the sounds that enter the ear are reasonably well preserved at the earliest stages of auditory processing, modulation of the mechanisms involved in auditory transduction may actually change the sounds as they first enter the system. As a result,

these altered sensations must necessarily propagate throughout the entire auditory system, or be actively corrected by the system at later stages of processing.

Here we investigate the influence of two multisensory processes that are well known to affect the auditory system – namely vision and eye movements – at the level of the auditory periphery. We conducted a series of three experiments to show that active mechanisms in the periphery respond to non-auditory stimuli and presumably instigate a cascade of events that either fundamentally alter the receptive properties of the ear and/or help the system integrate auditory signals with other sources of non-auditory information.

1.1 Non-auditory processes in the sub-cortical auditory system – a brief overview

There are multiple active mechanisms in the auditory periphery that might help to shape the flow of incoming information. Of particular interest are the outer hair cells (OHCs) of the cochlea, and middle ear muscles (MEMs). These mechanisms are both subject to descending control from more central regions of the brain and are capable of substantially altering incoming sounds.

While both mechanisms have previously been shown to be influenced by non-auditory processes, the extent to which they are influenced is a question that has been largely unaddressed.

The inferior colliculus (IC) is a brain region that lies more central to these mechanisms, but serves as a relay for nearly all ascending and descending auditory information (for review see Winer & Schreiner, 2005). Information traveling from the periphery to the cortex or vise-versa almost certainly passes through the IC. Importantly, it serves as a likely "point of entry" for many non-auditory inputs, as well as a possible conduit for multisensory signals from the cortex, which could then travel to the periphery and alter the function of the active mechanisms in and around the ear. Therefore, discussing the possibility of non-auditory processes occurring in the auditory periphery requires a brief review of the multisensory mega-highway in the auditory pathway.

The IC primarily sends ascending auditory information to the thalamus (e.g. Calford & Aitkin, 1983; Kudo & Niimi, 1980) which then proceeds toward the cortex, as well as the superior colliculus (SC) (Druga & Syka, 1984; Harting & Van Lieshout, 2000; Van Buskirk, 1983; S. Q. Zhang, Sun, & Jen, 1987). Prominent descending connections carry information back to the auditory brainstem (Huffman & Henson, 1990).

Converging anatomical and physiological evidence indicates that cells within the IC are sensitive to visual, oculomotor, and eye position, information as well as to signals relating to behavioral context and reward.

Numerous anatomical studies have established the existence visual innervation of the IC. Direct connections from the retina innervate the contralateral IC (mole-lemming: Herbin, Reperant, & Cooper, 1994; rat, monkey: Itaya & Van Hoesen, 1982; rat: Yamauchi & Yamadori, 1982; guinea pig, hamster, rat: A. B. Zhang, 1984), while inputs from visual cortex are sent to the ipsilateral IC (cat: Cooper & Young, 1976) and superior colliculus, a visually-responsive structure involved in programming saccadic eye movements (cat: Adams, 1980; rat: Coleman & Clerici, 1987; bat: Covey, Hall, & Kobler, 1987; barn owl: Hyde & Knudsen, 2000) (for review of saccade generation in the SC, see Gandhi & Katnani, 2011). Several physiological studies have established that there are cells in the IC whose auditory responses are modulated by a concurrent visual stimulus (Syka & Radil-Weiss, 1973; Tawil, Saade, Bitar, & Jabbur, 1983), or that are capable of responding directly to visual stimuli without an accompanying sound (Bulkin & Groh, 2012b; Mascetti & Strozzi, 1988; K. K. Porter, Metzger, & Groh, 2007).

While no studies have explicitly established the presence of strictly visual activity in the OHCs or MEMs, numerous studies have linked visual attention to the

OHCs. Specifically, attending to a visual stimulus tends to inhibit OHC response amplitude (de Boer & Thornton, 2007; Delano, Elgueda, Hamame, & Robles, 2007; Ferber-Viart, Duclaux, Collet, & Guyonnard, 1995; Froehlich, Collet, & Morgon, 1993; Meric & Collet, 1992, 1994; Meric, Micheyl, & Collet, 1996; Puel, Bonfils, & Pujol, 1988; D. W. Smith, Aouad, & Keil, 2012; D. W. Smith & Keil, 2015; Srinivasan et al., 2014; Srinivasan, Keil, Stratis, Woodruff Carr, & Smith, 2012). This suppression of activity is mediated by inhibitory inputs to the OHC population from the medial olivary complex (Guinan, 2006; Thier & Mock), which receives input from the IC and may also send projections, indirectly, to the MEMs (Mukerji, Windsor, & Lee, 2010). These are consistent with findings in the IC which show that anticipation, task engagement, and behavioral state all influence auditory processes (Metzger, Greene, Porter, & Groh, 2006; Nienhuis & Olds, 1978; Rinne et al., 2008; Ruth, Rosenfeld, Harris, & Birkel, 1974; A. Ryan & Miller, 1977; A. F. Ryan, Miller, Pfingst, & Martin, 1984), and are supported by anatomical inputs to the IC from regions typically associated with habitual (e.g. Yin & Knowlton, 2006) and motivated (e.g. Ono, Nishijo, & Nishino, 2000) behaviors (c.f. globus pallidus: Shammah-Lagnado, Alheid, & Heimer, 1996; Shinonaga, Takada, Ogawa-Meguro, Ikai, & Mizuno, 1992; Yasui, Kayahara, Kuga, & Nakano, 1990) (substantia nigra pars lateralis: Coleman & Clerici, 1987; ventral tegmental area: Herbert, Klepper, & Ostwald, 1997; basal nucleus of the amygdala: Hopkins & Holstege, 1978;

Marsh, Fuzessery, Grose, & Wenstrup, 2002; Moriizumi, Leduc-Cross, Wu, & Hattori, 1992; Yasui, Nakano, Kayahara, & Mizuno, 1991).

Closely associated with vision, as well as visual attention, is eye position. The eyes and ears necessarily receive visual and auditory spatial information, respectively, within a different frame of reference. Specifically, visual space is initially encoded based on where the image falls on the retina (a so-called eye-centered reference frame) while auditory space is calculated based on the position of the sound relative to the ears and the head (a head-centered reference frame). In species where the eyes are able to move to a substantial degree within the head (e.g. rhesus monkeys and cats but not rodents or barn owls), the visual and auditory reference frames are not fixed to each other. Because of this constantly changing relationship between reference frames, aligning visual and auditory space requires factoring in both the orbital position of the eyes and sound localization cues. Moreover, eye movements are closely associated with visual attention in the special domain, in that when making an eye movement to some location, that location is almost certainly the focus of visual attention.

Eye position signals are found regularly throughout the auditory system and related motor areas during both task-related and spontaneous fixations, and with or without the presence of a concurrent sound stimulus (Bulkin & Groh, 2012a; IC: Groh,

Trause, Underhill, Clark, & Inati, 2001; superior colliculus: Jay & Sparks, 1984; Lee & Groh, 2012; visual intraparietal sulcus: Mullette-Gillman, Cohen, & Groh, 2005; Mullette-Gillman, Cohen, & Groh, 2009; Populin, Tollin, & Yin, 2004; K. K. Porter, Metzger, & Groh, 2006; auditory cortex: Werner-Reiss, Kelly, Trause, Underhill, & Groh, 2003; Zwiers, Versnel, & Van Opstal, 2004). Not surprisingly, eye-movement related signals are also found distributed throughout the IC (Bulkin & Groh, 2012b) and may help support eye-position related activity. However, despite the physiological studies identifying eye position signals in the auditory system, the anatomical sources of these signals have yet to be identified. It is not clear whether such signals are a result of corollary discharge from oculomotor regions (e.g. the SC, or the fastigial nucleus of the cerebellum [Carpenter, 1959; Earle & Matzke, 1974]); somatosensory feedback from muscles controlling eye position, which could influence numerous auditory regions including the IC, auditory cortex, or cochlear nucleus (cuneate nuclei [J. D. Porter, 1986]; along, for instance, the opthalamic tract of the trigeminal nerve [Steinbach, 1987]; or from the somatosensory cortex [M. Zhang, Wang, & Goldberg, 2008]); or some combination of corollary discharge and somatosensory signals. Interestingly, MEM activity has been associated with rapid eye movement (REM) sleep, and may occur with or without associated eye movements (De Gennaro & Ferrara, 2000; De Gennaro, Ferrara, Urbani, & Bertini, 2000; Dewson, Dement, & Simmons, 1965; Pessah & Roffwarg, 1972). This suggests that these eye-position related signals may originate, or at

least begin to do so, in the auditory periphery. Regardless of where they first enter the system though, they may easily ramify throughout.

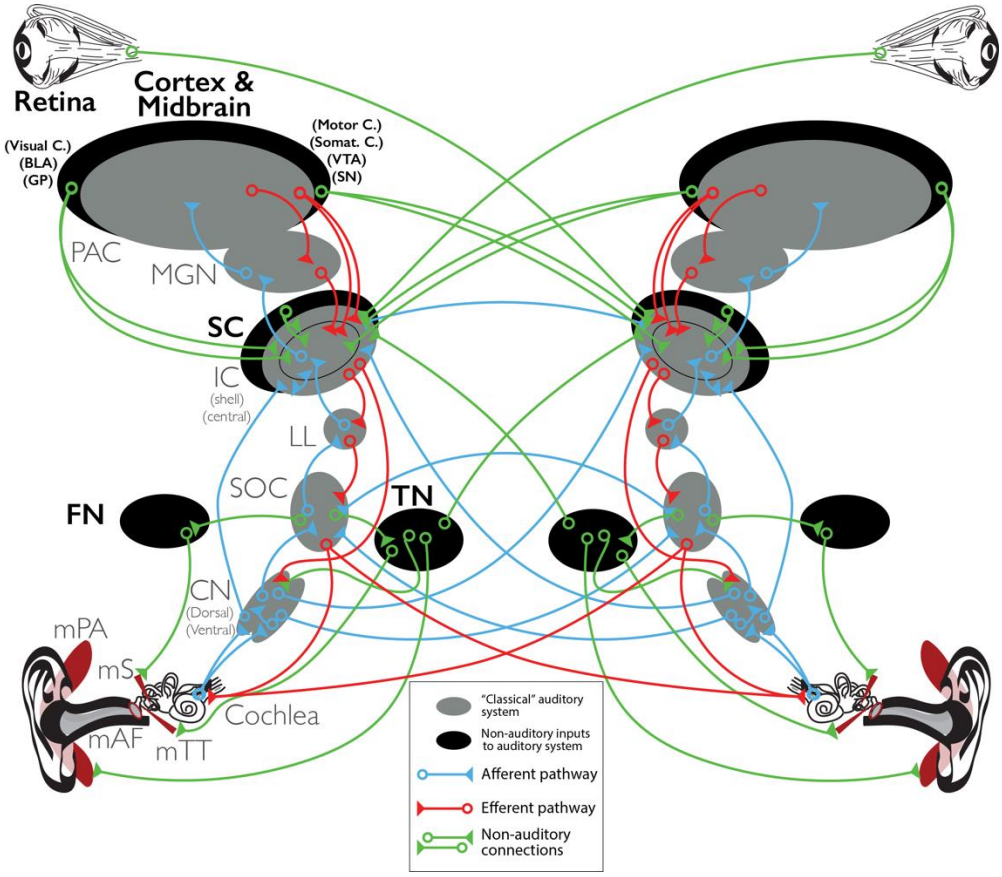

Figure 1: Connectivity within the auditory system, particularly with regards to non-auditory inputs

The auditory system has prominent afferent (cyan) and efferent (red) subdivisions, both of which span the entire auditory pathway and provide multiple possible routes for information to get from one subdivision to another. There are also numerous non-auditory (green) inputs to the

system which carry information from regions in the cortex and brainstem involved in movement, proprioception, and behavior, including both vision and eye movements.

In particular, visual regions (visual cortex, SC, and retina), eye-movement related regions (SC, somatosensory cortex, TN), and behavioral regions (BLA, GP, VTA, SN) send non-auditory inputs to the IC, which can then be sent to peripheral mechanisms (Cochlea [OHCs, in particular], mTT, mS, mAF, mPAG).

The circuitry of the lower auditory system helps to tune the active mechanisms in response to sound input. Sounds from one ear are passed into the SOC where crossed and uncrossed connections provide feedback directly to the OHCs of the cochlea, or indirectly to the musculature in and around the ear. The motor neurons of the stapedius muscle reside near the facial nerve nucleus while those of the tensor tympani are near the trigeminal nerve nucleus alongside the motor neurons of the post-auricular muscles near the pinnae. Both pools of motor neurons receive inputs from the SOC and potentially other sources. Additionally, the muscles of the annulous fibrosus surrounding the tympanic membrane may exhibit control over the membrane, but little is known about the role or innervation of these muscles. This local network also receives input from higher brain regions that further refines these peripheral filter mechanisms prior to or concurrent with sound onset; these connections may carry information regarding eye movements and attentional status.

BLA – basolateral amygdala; BM – basilar membrane; CN – cochlear nucleus; FN – facial nerve nucleus; GP – globus pallidus; IC – inferior colliculus; LL – lateral lemniscus; MEM – middle ear muscles; MGN – medial geniculate nucleus; mAF – annulus fibrosus muscles; mPA – post-auricular muscles; mS – stapedius muscle; mTT – tensor tympani muscle; OHC – outer hair cell; PAC – primary auditory cortex; SC – superior colliculus; SN – substantia nigra; SOC – superior olivary complex; TM – tectorial membrane; TN – trigeminal nucleus; VTA – ventral tegmental area

1.2 Summary of, and rationale for, experiments run

Because of the convergence of possible non-auditory inputs from more central regions in addition to potentially related attentional and eye-movement related activity, we developed a series of three experiments to study the effects of eye-movements and vision on the auditory periphery. Both MEMs and OHCs are capable of vibrating the eardrum via mechanical interactions of structures in the ear, and these vibrations can be recorded as acoustic signals by a microphone set into the ear canal. All three experiments take advantage of this phenomenon, which we use as a measure of the collective response of all mechanisms, including the OHCs and MEMs, that may affect

these acoustic recordings. Here we briefly summarize each experiment, including the specific rationale, methodology, and findings of each.

1.2.1 The eardrum moves when the eyes move: A multisensory effect on the mechanics of hearing

Interactions between sensory pathways are known to occur in the brain, but where they first occur is uncertain. Here we show a novel multisensory interaction occurring at the level of the eardrum. Ear canal microphone measurements in humans (n=19 ears in 16 subjects) and monkeys (n=5 ears in 3 subjects) performing a saccadic eye movement task to visual targets indicated that the eardrum moves in conjunction with the eye movement. The eardrum motion was oscillatory and began as early as 10 ms prior to saccade onset in humans or at the time of saccade onset in monkeys. It lasted through saccade offset in both species. The eardrum movement, which we dub the Oscillatory Saccadic Eardrum Associated Response or OSEAR, was seen in 16 of 19 human ears and all 5 monkey ears. The amplitude and phase of the oscillation depended on the direction and horizontal amplitude of the saccade and occurred in the absence of any externally delivered sounds. We conclude that OSEARs create a novel eye movement-related binaural cue that may aid the brain in evaluating the relationship between visual and auditory stimulus locations as the eyes move.

1.2.2 Anticipation of an upcoming sound alters acoustic output of the ear

Where and how different sensory pathways converge in the brain is presently unknown, and of great relevance for understanding how the senses communicate. Vision is known to exert multiple effects on auditory perception, including sound localization and speech recognition, and vision in the absence of auditory cues can modulate the activity of peripheral mechanisms in the auditory system. It is known that attending visual stimuli can inhibit peripheral responses to incoming sounds, and our lab has recently demonstrated that eye movements – which are necessary to synchronize auditory and visual space – influence peripheral auditory processes. Here, we demonstrate that the periphery is also sensitive to visual stimuli that are predictive of upcoming sounds when eye position is held constant. By measuring the acoustic pressure in the ear canal during an audiovisual cueing task, we observe a dynamic change to pressure levels in the ear when a sound is expected versus when one is not. Specifically, we observe a drop in frequency power and variability within the behaviorally important frequency range of 0.1 to 4 kHz around the time when the sound is expected to occur; this is in contract to a slight increase in power at both lower and higher frequencies. We suspect these changes are due to the contraction of the tensor tympani muscle in the middle ear and help to minimize system noise in order to allow the anticipated signal to more readily pass into the cochlea.

1.2.3 McGurk effect in the auditory periphery: The influence of mismatched audiovisual speech cues on active mechanisms in the ear

Speech is typically thought of as a form of auditory communication, but viewing the movements of a speaker's lips can inform and even alter the auditory percept of the speech sounds. Where and how the brain combines visual and auditory cues in speech remains an open question. Although interpretation of audiovisual speech cues is most often associated with cortical processes, descending connections in the auditory system make it possible in principle for visual information to modify auditory processes early in the auditory pathway. We investigated whether visual speech cues can influence the peripheral auditory system using the well-known audiovisual speech illusion known as the McGurk effect. Here, we show that seeing a speaker say a syllable that is incongruent with the accompanying audio can alter the response patterns of the auditory periphery during the most relevant moments in the speech stream. These visually influenced changes may contribute to the altered percept of the speech sound.

1.3 Goal and intent of research project

This series of experiments is meant to be a pioneering study of multisensory processes at the auditory periphery. The studies are designed to test well-known and accepted principles of sensory integration within the auditory system and synthesize them with known behavior of active mechanisms in the auditory periphery in order to test whether the peripheral receptors are subject to complex non-auditory processes.

These studies are also meant to expand on the technique of recording acoustic signals from the ear canal as a simple, effective, inexpensive, and non-invasive method for multisensory and other comparative studies. This technique was originally used to study the activity of the MEMs during the mid-1900s (e.g. Weiss, Mundie, Cashin, & Shinabarger, 1962), but with the discovery of otoacoustic emissions as generated by the OHCs by Kemp (Kemp, 1978, 1979), the MEMs and other processes in the ear are now typically ignored or specifically removed from the recordings. However, there may be cases where no individual mechanism is necessarily implicated in some process; acoustic measurements allow for a more generalized study of all possible peripheral mechanisms without making assumptions about which mechanisms can or cannot be involved in a particular behavior. While this technique is inherently limited in its inability to specifically isolate individual active mechanisms, these various mechanisms often have

signature activity patterns that can provide clues as to the mechanisms involved in a particular recording. These recordings can, therefore, provide valuable information for more detailed probes of specific mechanisms while simultaneously remaining general enough to detect differences in activity across test parameters whose sources are otherwise unknown.

Therefore, it is our goal that this set of experiments both expands our knowledge of peripheral auditory function with specific regard to non-auditory influences on the system, as well as our ability to conduct further such study in the near future.

2. The eardrum moves when the eyes move: A multisensory effect on the mechanics of hearing

The visual system aids auditory processing. For example, visual lip reading cues facilitate auditory speech comprehension and can create illusions such as the McGurk and ventriloquism effects (Campbell, 2008; L. Chen & Vroomen, 2013; Kopco, Lin, Shinn-Cunningham, & Groh, 2009; McGurk & MacDonald, 1976; Recanzone, 1998). Historically, visual and auditory signals were presumed to remain segregated through the initial stages of their respective pathways and only converge later in association areas such as parietal cortex or the superior colliculus (historical review on association cortex: Chow & Hutt, 1953; Drager & Hubel, 1975; Gordon, 1973; Mast & Chung, 1973; Paula-Barbosa & Sousa-Pinto, 1973; SC, e.g.: Wickelgren, 1971). More recent evidence shows that visual information is present throughout much of the auditory system, particularly in auditory cortex (e.g. Brosch, Selezneva, & Scheich, 2005; Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Schroeder & Foxe, 2002), and inferior colliculus (Bulkin & Groh, 2012b; Gutfreund, Zheng, & Knudsen, 2002; Mascetti & Strozzi, 1988; K. K. Porter et al., 2007), while influences related to visual attention have also been found in the auditory cortex(e.g. Molloy, Griffiths, Chait, & Lavie, 2015), inferior colliculus (Metzger et al., 2006), and even the outer hair cells in the cochlea (c.f. de Boer &

Thornton, 2007; Delano et al., 2007; D. W. Smith et al., 2012; D. W. Smith & Keil, 2015; Srinivasan et al., 2014; Srinivasan et al., 2012).

Eye-movement and position related signals, which are thought to be an essential element of communication between the visual and auditory pathways in the spatial domain, and are closely associated to visual attention, are known to collocate in many of these same brain regions (Bulkin & Groh, 2012a, 2012b; Fu et al., 2004; Groh et al., 2001; Lee & Groh, 2009; Maier & Groh, 2010; Metzger, Kelly, & Groh, 2004; Mullette-Gillman et al., 2005, 2009; K. K. Porter et al., 2006; Werner-Reiss et al., 2003; Zwiers et al., 2004). Without information about the position of the eyes in the orbits, the brain would be unable to relate information about sound location derived from head-centered binaural difference cues to information about visual stimulus location derived from eye-centered sites of retinal activation (Groh & Sparks, 1992). Because of their importance and intimate involvement in visual- and attention-related influences found throughout the auditory system from cochlea to cortex, it is possible that eye-movements are integrated into the pathway from the time when sounds first enter the ear.

The auditory periphery possesses at least two means of tailoring inputs to the system in response to descending neural control: (1) The middle ear muscles (MEMs) – the stapedius and tensor tympani – attach to the ossicles that connect the eardrum to the

oval window of the cochlea. Contracting these muscles tugs on the ossicular chain, generating motion of the eardrum (figure 2, blue inset). (2) Within the cochlea, the outer hair cells (OHCs) are mechanically active and amplify the motion of the basilar membrane's response to sounds. They can also alter the motion of the membrane in response to descending neural commands (figure 2, purple inset). Both the MEMs and OHCs are subject to descending control by signals that may originate in various parts of the brain but which ultimately pass through the superior olive (see for review: Guinan, 2006 [OHCs]; Mukerji et al., 2010 [MEMs]), and both mechanisms affect eardrum motion through mechanical coupling of the eardrum, ossicles, and inner ear. Neural control over the auditory periphery can therefore be measured by placing a microphone in the ear canal and recording sounds resulting from centrally-mediated eardrum motion (e.g. measurement of otoacoustic emissions). We used this technique to study the effect of eye movements on the peripheral mechanisms – specifically the OHCs and MEMs – controlling the eardrum both in response to and in the absence of sounds.

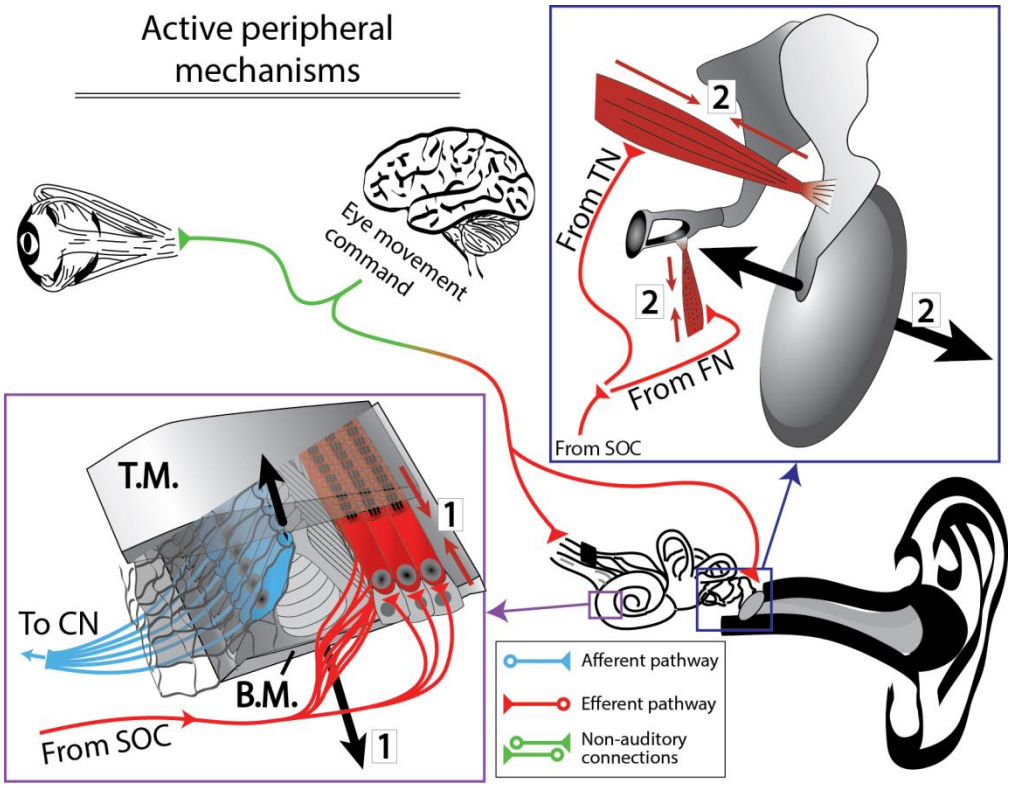

Figure 2: The auditory periphery possesses multiple active mechanisms capable of altering its responsiveness based on commands originating in the brain.

Purple inset: When the outer hair cells (OHCs) of the cochlea (red cells) expand or contract, they move the basilar membrane, which through the fluid/mechanical hydro coupling between the basilar membrane, oval window, and ossicular chain results in movements of the eardrum [1]. This reverse traveling wave can be measured with a microphone in the ear canal as an otoacoustic emission. These cells are innervated by neurons originating in the superior olive bilaterally. Blue inset: Contraction of the middle-ear muscles pulls on the ossicles which in turn move the eardrum (or tympanic membrane) [2]. These muscle fibers are innervated by motor neurons that reside near the facial and trigeminal nerve nuclei, which both receive input from the superior olive bilaterally. Both the OHCs and MEMs receive direct or indirect input from the

SOC, and movements of both mechanisms cause eardrum motion that can be recorded with a microphone set in the ear canal.

BM – basilar membrane; CN – cochlear nucleus; FN – facial nerve nucleus; MEM – middle ear muscles; OHC – outer hair cell; SOC – superior olivary complex; TM – tectorial membrane; TN – trigeminal nucleus

2.1 Results

Sixteen humans performed a task involving saccades to visual targets varying in horizontal position (figure 3a). Half of the trials were silent whereas the other half incorporated a series of task-irrelevant clicks presented before, during, and at two time points after the visually-guided saccade (figure 3b). This design allowed us to compare the effects of more central neural control over the auditory periphery in silence and in the presence of the type of brief sounds often used to elicit otoacoustic emissions.

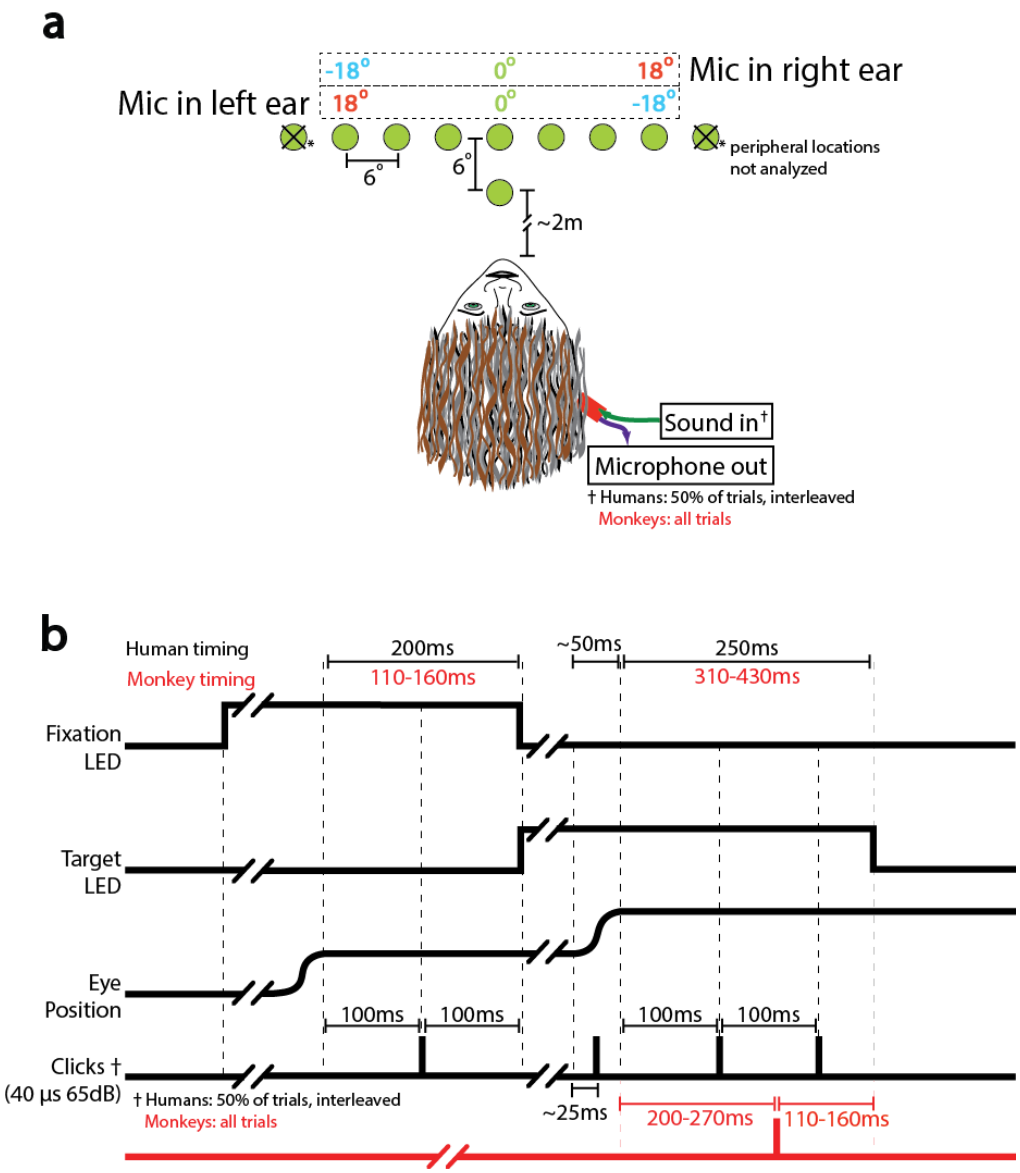

Figure 3: Spatial (a) and temporal (b) elements of the eye movement task.

On each trial, a subject fixated on a central LED, then made a saccade to one of nine target LEDs without moving his/her head. Target locations spanned -24º to +24º in the horizontal dimension and were located 6º above the fixation LED. The most peripheral targets (±24º) were included on 4.5% of trials to improve performance at the ±18º positions but were

excluded from subsequent analyses (see Methods); all other target locations were equally likely at 13% each. For human subjects, half of the trials were silent and the other half (which were randomly interleaved) involved brief clicks played at four times: during the initial fixation, during the saccade, and at 100ms and 200ms after target fixation was obtained. Monkey subjects performed a hybrid trial type with some minor difference in task timing (marked in red): all trials were silent up through the saccade, followed by a single click 200-270 ms after saccade offset (i.e. at roughly the same time as the second post-saccadic click for the human with-sound trial type; red click trace). Recordings were made within the ear canal via a microphone set into a custom-fit ear bud and calibrated at the start of each session. The full duration of each trial was recorded.

2.1.1 Saccades move eardrum in silence

We found that the eardrum moved when the eyes moved, in the absence of any externally delivered sounds. Figure 4 shows the average eye position (figure 4a) and velocity (figure 4b) as a function of time for each target location for the human subjects on trials with no sound stimulus. The corresponding average microphone readings are aligned on saccade onset and color-coded for saccade direction and amplitude (Figure 4c). The microphone readings oscillate time-locked to movement onset with a phase that depends on saccade direction; the spectral power is strongest at 20-40Hz with an amplitude of 3.36 mPa (44.5 dB) greater than the pre-OSEAR baseline (mean amplitude at -25 to -15 ms) and minimal activity above 60Hz (figure 5). When the eyes moved

toward a visual target contralateral to the ear being recorded, the microphone trace deflects slightly upward beginning about 10 ms prior to eye movement onset, followed by a more substantial downward deflection at about 5 ms after the onset of the eye movement. The microphone trace completes 2-3 cycles in total before returning to baseline at about 60 ms after the onset of the eye movement, which approximately corresponds to the conclusion of the eye movement. For saccades in the opposite (ipsilateral) direction, the microphone traces initially deflect downward and continued following a similar trajectory but opposite in sign. Figure 4d shows the same microphone data but separates out each pair of traces associated with eye movements of the same amplitude but to opposite hemifields, allowing the inclusion of the standard error of recordings (shaded areas around each trace).

To obtain an overall portrait of the statistical significance of the relationship between eye movements and the observed microphone signal across time, we calculated a regression of microphone voltage vs. saccade target location for each 0.04 ms sample from 25 ms before to 100 ms after saccade onset. Since this involved many repeated statistical tests, we compared the real results with a Monte Carlo simulation in which we ran the same analysis but scrambled the relationship between each trial and its true saccade target location (see methods section for details). As shown in figure 4e, the slope of the regression involving the real data (gold trace) oscillates during the saccade period,

beginning 9 ms prior to and continuing until about 60 ms after saccade onset, matching the oscillations evident in the data in figure 4c. In contrast, the scrambled data (gray trace) is flat during this period. Similarly, both the percent of significant subjects at each time point ($p<0.05$) and effect size (R^2 values) of the real data deviate from the scrambled baseline in a similar but slightly longer timeframe, dropping back to the scrambled data baseline at 75 ms after saccade onset (figure 6).

The overall amplitude of the oscillation was slightly larger for saccades directed contralateral to the recorded ear than for ipsilateral saccades. To evaluate this statistically, we calculated the root mean squared (RMS) amplitude of the microphone signal from 10 ms before saccade onset to 75 ms after, or approximately the duration of the significant region of our oscillatory activity (figure 6). The contralateral vs. ipsilateral difference was statistically significant (t-test, $p<0.05$; figure 4f). Within a hemifield, microphone RMS amplitude varied monotonically with eccentricity (figure 3g, regression of microphone RMS vs. locations L = {0, 6, 12, 18} or L = {-18, -12, -6, 0} significant $p<0.05$).

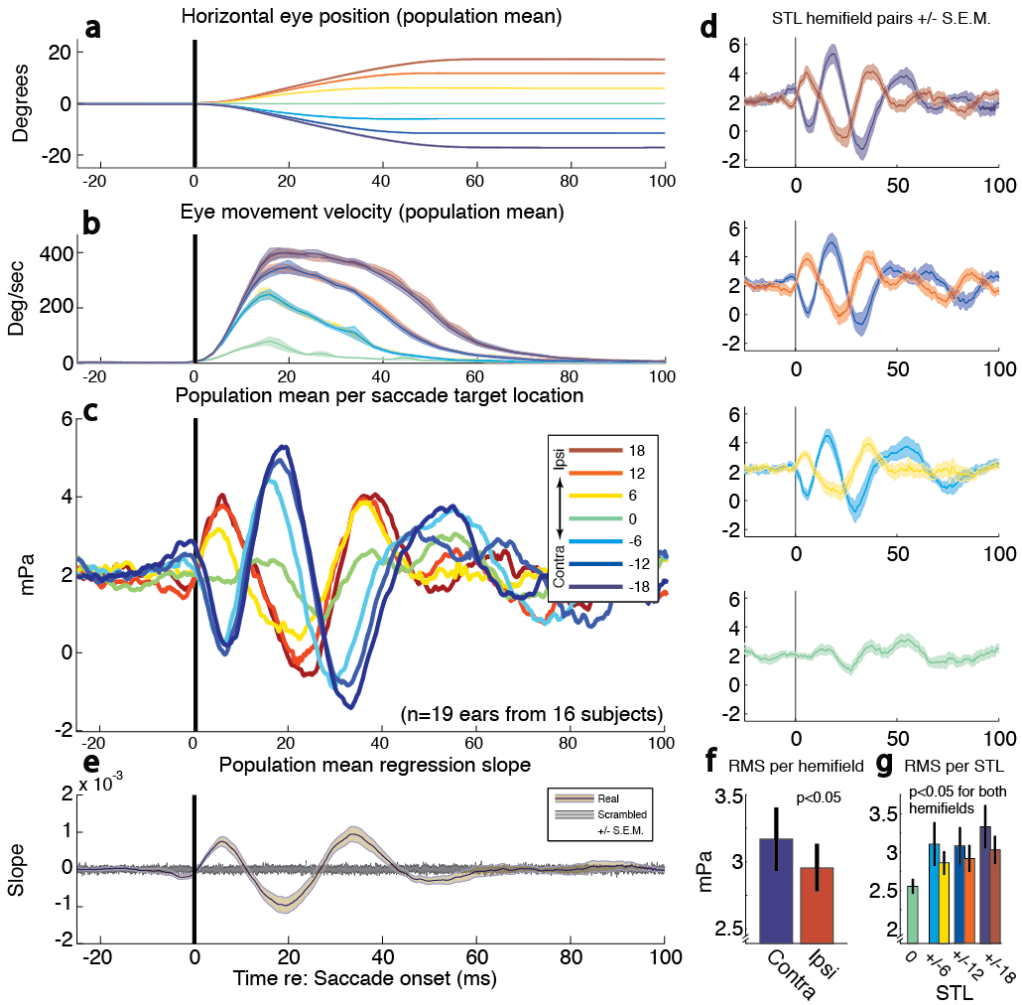

Figure 4: Eye movements and associated eardrum movements in the absence of a sound stimulus as measured with a microphone.

(a-b) Eye positon and velocity as a function of time, aligned on saccade onset; colors indicate saccade target locations (STL) from ipsilateral (red) to contralateral (blue). (c-d) Microphone recordings in the ear canal, aligned on saccade onset and color coded to match the associated eye movement. Raw microphone voltages for each STL were averaged for each subject, and these individual subject averages were converted to pressure and combined to form the group average shown in panel c. The data from panel c are shown again in the subpanels of panel d for

each mirror image pair of target locations with their corresponding standard errors (shaded colors). (e) Average slope of a regression of microphone reading vs. STL for each time point for real (gold) vs. scrambled (gray) data. Scrambling involved reassignment of each trial to a randomly chosen saccade target location. This analysis was conducted separately for each of the 19 human ears and the results were pooled by averaging the observed slopes across ear-subjects; shading indicates ± standard error. (f-g) Root mean squared (RMS) amplitude from 10 ms before to 60 ms after saccade onset, by hemifield (panel f), and individual target locations (panel g). RMS amplitudes were calculated individually for each subject for all STLs; for panel f, STLs within the same hemifield were averaged (ipsilateral = {6, 12, 18}; contralateral = {-18, -12, -6}). The amplitude of the eye movement-related eardrum oscillation was significantly larger for contra vs. ipsi eye movements (t-test, $p<0.05$). Within hemifields, (i.e. STL = {0, 6, 12, 18} or {-18, -12, 6, 0}; bars are color coded to STL as in panels a-d), regressions indicated that the RMS amplitudes varied significantly with STL ($p<0.05$). All data included here involved correctly performed trials meeting noise-rejection inclusion criteria as described in the methods.

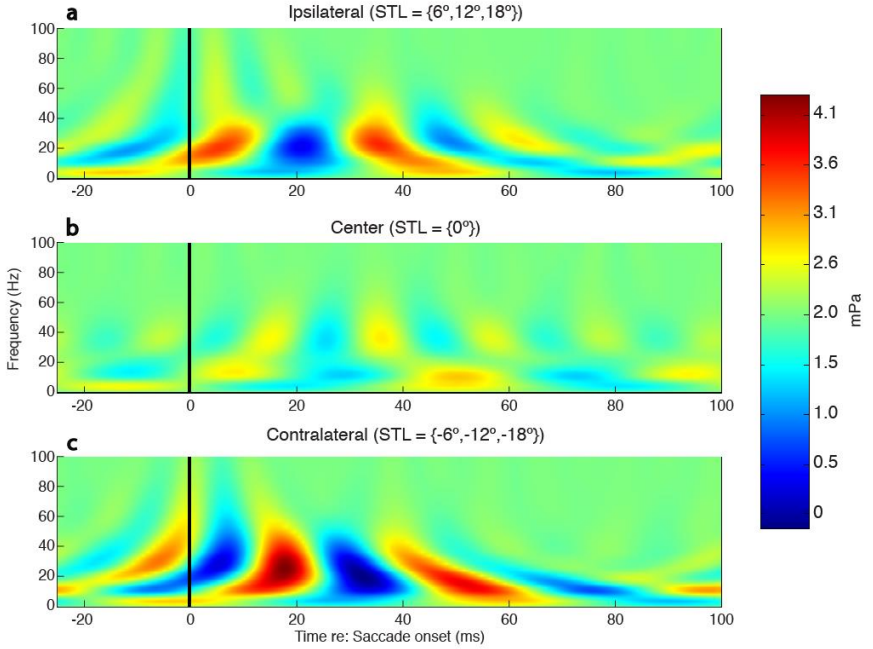

Figure 5: Frequency content over time as a function of saccade target location hemisphere.

All population mean traces within the same hemisphere – (a) ipsilateral, (b) center, and (c) contralateral – were combined into a grand mean and deconstructed with wavelet analysis (Stanford Wavelabs package for Matlab). Note the concentration of frequency power from approximately 15 to 60 Hz, with a peak near 20 Hz. These frequency ranges are consistent with previously observed oscillatory frequencies observed in both MEMs.

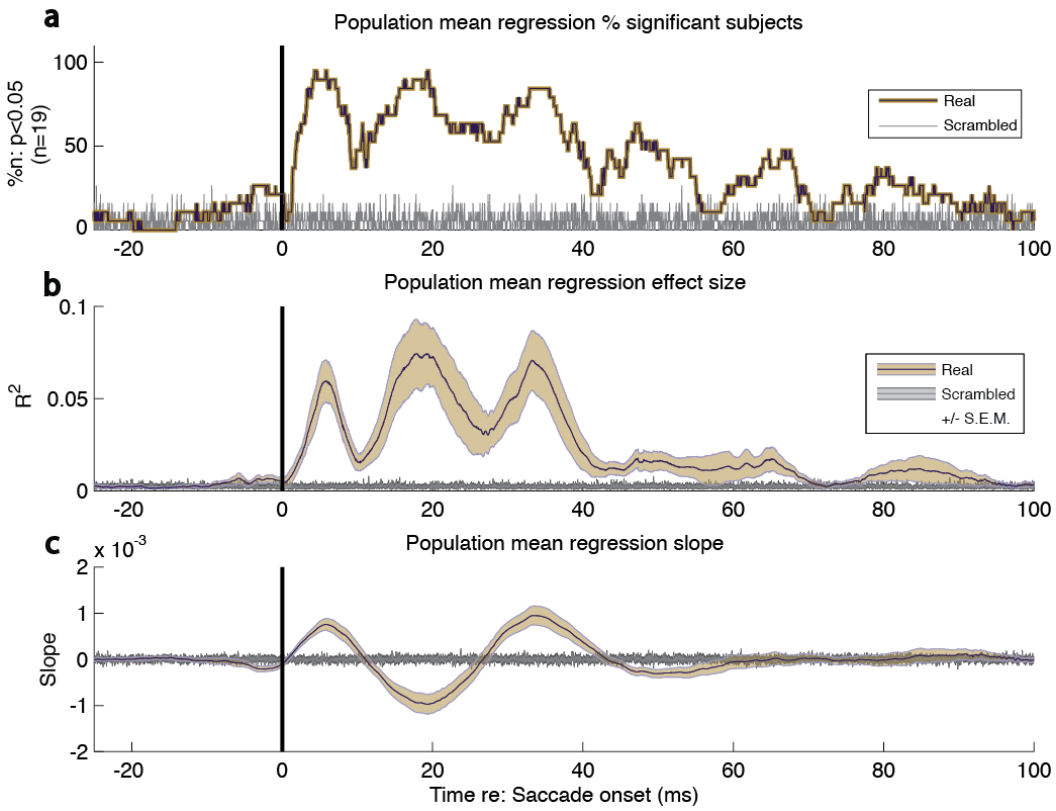

Figure 6: Full regression results for human subjects.

(a) Percent of subject ears showing p <0.05 for the corresponding time point. (b) Proportion of the variance accounted for by the regression (R^2). (c) Slope (same as figure 4e). Methods discussed in detail in methods.

The oscillatory saccadic eardrum associated response (OSEAR), was apparent in most individual subjects, with 16 out of 19 recorded ears (from 16 human participants) showing a significant effect (figure 7a and b; $p<0.05$ for ANOVA aligned on first peak in population mean: 4.5 to 6.5 ms post saccade onset). All three subjects for whom both ears were recorded showed significant OSEARs in both ears, although the magnitude of

the OSEAR was generally different (note the different y-axis scales for the two ears of subjects HMJ050 and HWJ070). One of the three showed a slight individual idiosyncrasy in the timing and wave form of the OSEAR, which was present in both ears (HMJ050, figure 7b middle row, black and gray arrowheads; see figure caption for explanation). The three non-significant ears are shown in figure 7c; some evidence of OSEARs are apparent even in these subjects.

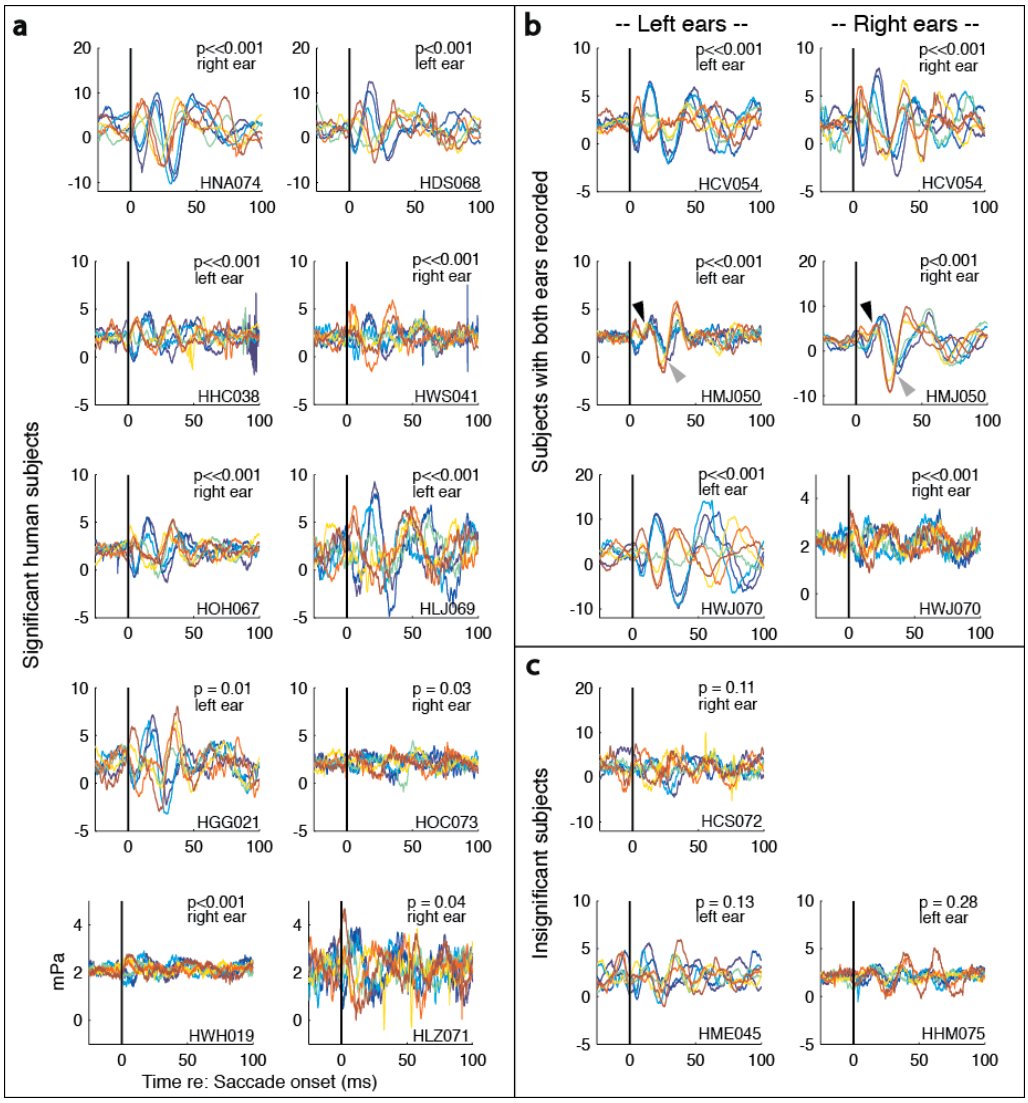

Figure 7: Mean traces for all individual human subjects' ears.

Significance was determined by an ANOVA with saccade target location as the independent factor and the mean microphone value for each trial within the window from 4.5 to 6.5 ms after saccade onset as the dependent factor. This time window was determined based on the location of the initial peak in the population mean traces (figure 4c). ANOVA testing was calculated on all trials included for that subject (i.e. correct, low noise trials; see methods). (a) All

subjects showing significant (ANOVA, $p<0.05$) effects in the one ear recorded. Recording amplitudes varied per subject, and y-axis scales are adjusted accordingly (top row; middle three rows; bottom row). Subjects within the same y-axis scale are arranged according to significance from left to right, top to bottom. (b) Three subjects for whom both ears were recorded. Note the change in y-axis scale between the two ears for subject HMJ050 (row 2) and HWJ070 (row 3). Two examples of notable individual idiosyncrasies preserved between both ears are marked for HMJ050: ipsilateral traces reset phase about half-way through first cycle and realign with contralateral traces (black arrowhead), then the contralateral traces lag behind the ipsilateral (gray arrowheads). (c) Three subjects whose ears did not reveal significant OSEARs by our measure. Note that these subjects still had some oscillatory activity as seen in our significant subjects, but the relationship between the microphone signal and saccade target location at the chosen time did not reach statistical significance.

As noted earlier, eye movements are known to affect auditory processing in several areas of the auditory pathway (Bulkin & Groh, 2012b; Fu et al., 2004; Groh et al., 2001; Lee & Groh, 2009; Maier & Groh, 2010; Metzger, Mullette-Gillman, Underhill, Cohen, & Groh, 2004; Mullette-Gillman et al., 2005, 2009; K. K. Porter et al., 2006; Werner-Reiss et al., 2003; Zwiers et al., 2004). Because this previous work was conducted using non-human primate subjects, we sought to determine whether OSEARs also occur in monkeys. We tested the ears of rhesus monkeys (*macaca mullata*; n=5 ears in 3

monkeys) performing a paradigm involving a trial type that was a hybrid of those used for the humans. The hybrid trial type was silent until after the eye movement, after which a single click with a 200-270 ms variable delay. OSEARs were observed with roughly similar timings, both at the population level (figure 8) and individual subject level (figure 10; $p<<0.001$ for all ears). The time-wise regression analysis suggests that the monkey OSEAR begins weakly at the time of the saccade and reaches a robust level about 10 ms later (figure 8e; percent significant and effect size: 2.8). RMS analyses were not significant for the monkey subjects.

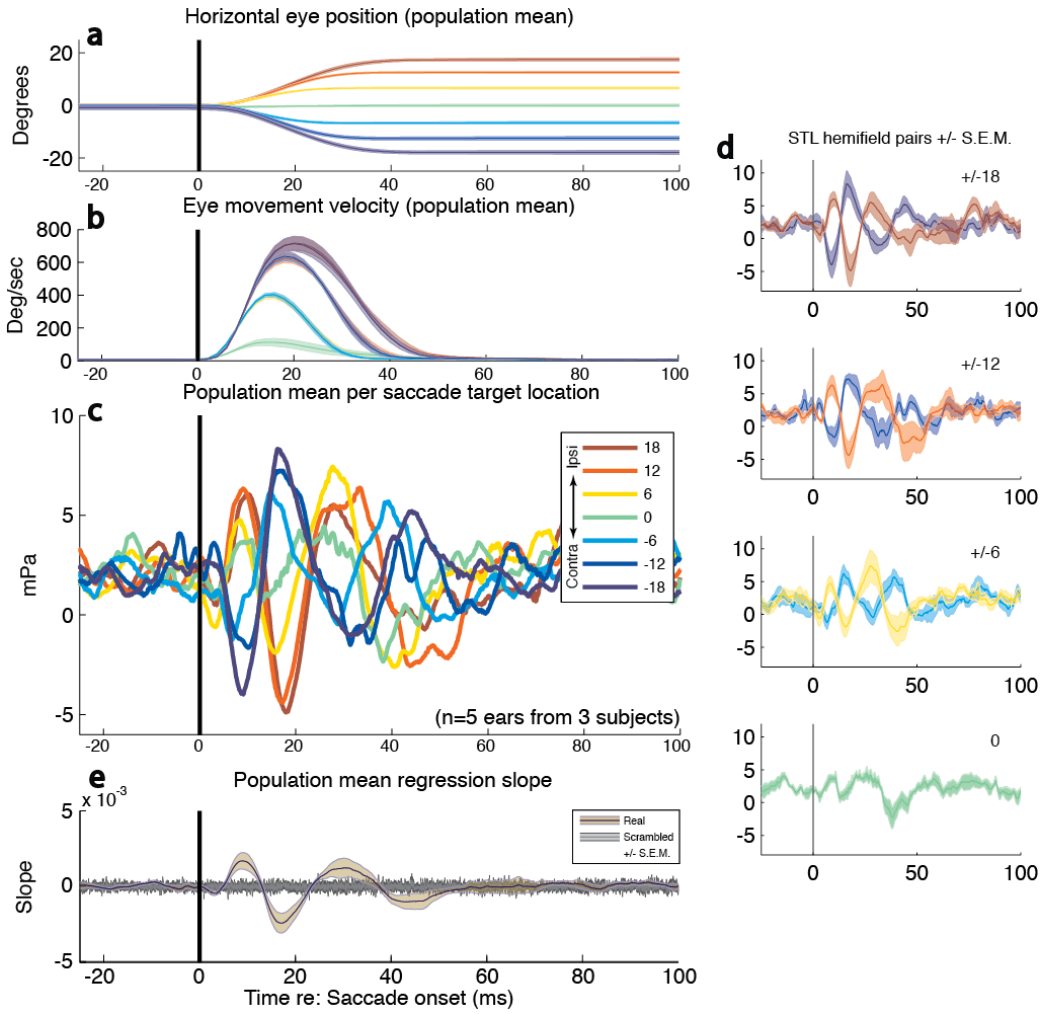

Figure 8: Results for monkey population (n=5 ears from 3 animals).

All sub panels a-e were calculated the same as in figure 4 (human data). Panels f-g were excluded as these data did not reach significance for the relatively small number of monkey ears recorded. There were minor differences in the paradigm used for humans and for monkeys (see methods for complete details on difference), most importantly that: 1) eye tracking was done with a scleral eye coil; 2) all trials involved only one click, presented at 200-270 ms after target fixation was obtained; and 3) the number of trials per session and per subject were substantially different

than the human paradigm, with monkeys MNN012 and MYY002 performing about 4000 and 3000 trials per ear respectively over four sessions while monkey MHH003 only performed 200 correct trials in total. Despite having so few trials, data from MHH003 are consistent with the rest of the human and monkey population (see figure 10) and show that these effects can be robust even when only a few trials are collected.

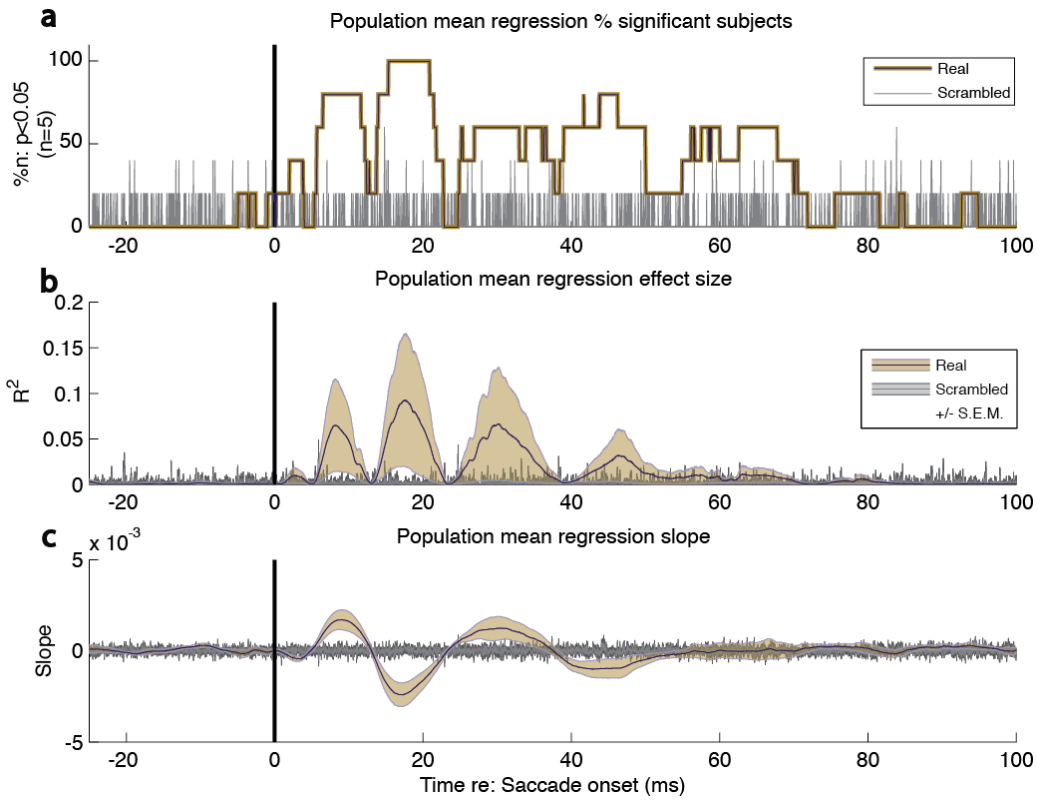

Figure 9: Expanded regression results for monkey population.

Percent significant subjects (a), R^2 (b), and slope (c; the same as figure 8e). Methods discussed in detail in methods.

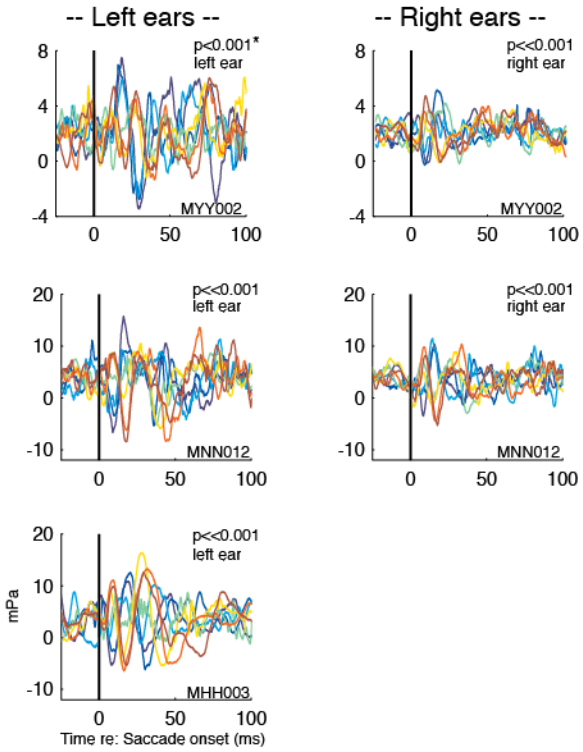

Figure 10: Individual subject results for all monkey ears.

*Analysis details are the same as individual human subjects, but the timing of the first peak is slightly delayed relative to the human data. The ANOVA measurement is therefore taken at the monkey's first peak, from 8 to 10ms post saccade onset. * Note that the left ear of subject MYY002 had a relatively delayed effect compared to the other monkey ears such that the first peak of activity in this ear was reasonably well aligned with the second peak for the rest of the group. Accordingly, this ear is not significant at the same time window (8 to 10ms: p=0.93), but if the ANOVA window is defined with respect to the second population peak at 16 to 18 ms, the ANOVA is significant (p<0.001).*

2.1.2 OSEARs are not due to electrical artifact

We next considered possible sources of electrical artifact that could have contributed to these observations. In particular, if the microphone's circuitry acted in part as an antenna, it could have been influenced by sources of eye-movement related electrical signals in either our eye movement measurement system in monkeys (scleral eye coils) and/or some other component of the experimental environment, electrooculographic (EOG) signals which result from the electrical dipole of the eye, or myogenic artifact such as electromyographic (EMG) signals originating from the extraocular or various facial muscles, including the auricular muscles. If such artifacts contributed to our measurements, they should continue to do so when the microphone was acoustically plugged without additional electrical shielding. Accordingly, we selected four of the subjects who showed a significant effect of eye movements on the microphone signal (figure 11a, left column) and repeated the test while the acoustic port of the microphone was plugged. The eye movement-related effects were no longer evident ($p>0.1$) when the microphone was acoustically occluded (figure 11a, right column). This shows that the eye movement-related change in the microphone signal stems from its capacity to measure acoustic signals rather than electrical artifacts. Additionally, OSEARs were not observed when the microphone was placed in a 1mL syringe (the approximate volume of the ear canal) and positioned directly behind the

pinna of a human subject (t-test, p=0.43; figure 12). In this configuration, any electrical contamination should be similar to that in the regular experiment; seeing none therefore supports the interpretation that the regular ear-canal microphone measurements are detecting eardrum motion and not electrical artifact.

2.1.3 OSEARs in the presence of loud contralateral noise

To verify that brain-controlled active mechanisms in the ear are the source of OSEARs, we sought to determine if OSEARs change when those active mechanisms are manipulated. Loud sounds in either ear are known to trigger bilateral OHC and MEM activation in order to adjust the gain of incoming signals and protect delicate inner-ear structures (c.f. Gelfand, 1984 [MEMs]; D. W. Smith & Keil, 2015 [OHCs]; Zhao & Dhar, 2010). This reflex reduces the dynamic range of both mechanisms and should attenuate the amplitude of eardrum motion. This suggests that OSEARs should change, and likely be attenuated, when loud sounds are present. We tested this theory by repeating the paradigm for seven subjects with significant OSEARs (figure 11b, left column) while playing a 90dB SPL white noise burst into the contralateral ear for the duration of each trial (figure 11b, right column). All subjects showed a reduction in RMS amplitude of the OSEAR (figure 11b, text column), and in 3 of the 7 cases, the ANOVA measure of

OSEAR presence was no longer significant. These data demonstrate that the OSEAR is influenced by the crossed/bilateral acoustic reflex mediated by brain circuitry.

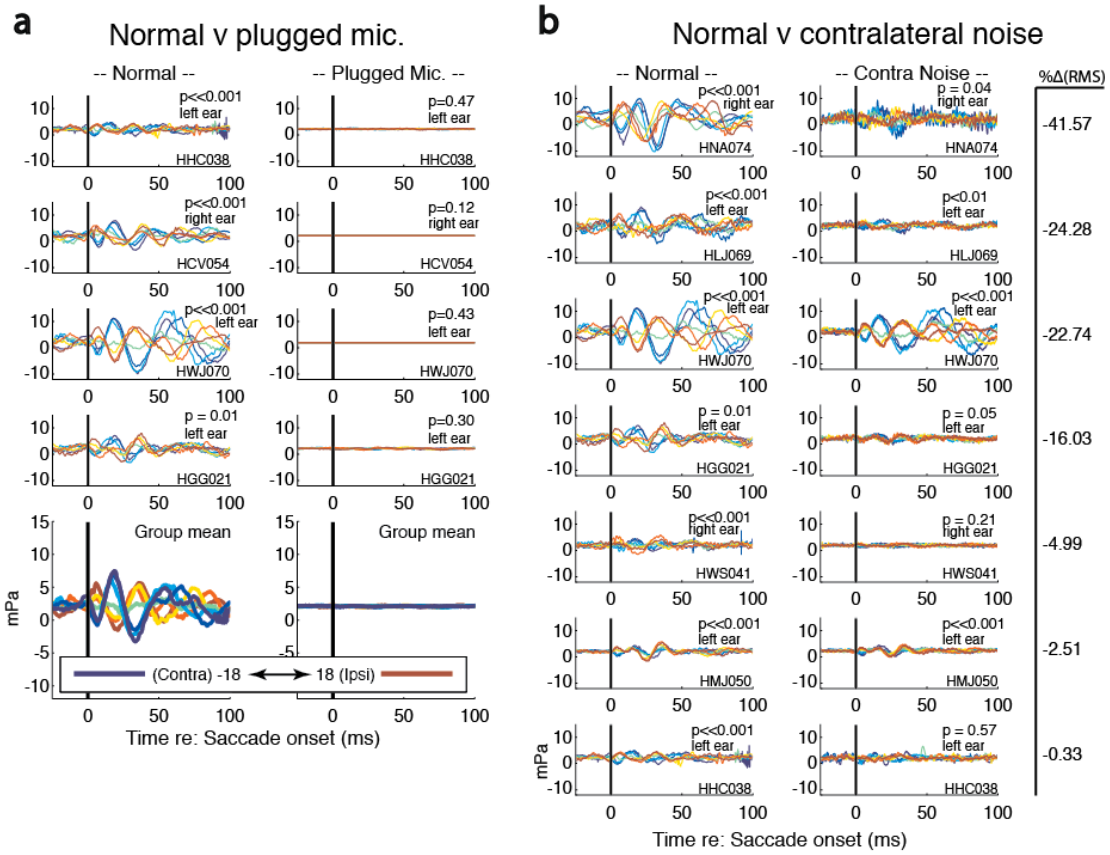

Figure 11: Control measures as compared to same subject in normal paradigm.

OSEARs are not observed when the microphone is plugged, eliminating acoustic but not electrical contributions to the microphone signal (a), and are modified by conditions known to trigger changes in brain-mediated peripheral auditory gain control (b). (a) Normal (left column) recordings versus recordings taken from the same ear but with the microphone port plugged (right column). The plugged microphone sessions were run exactly as the normal sessions except that after calibration, the microphone was removed, set into a closed earbud, and re-inserted into

the ear canal. The closed earbud acoustically shielded the microphone from the rest of the ear canal. (b) Normal (left column) recordings versus recordings taken from the same ear while white noise (uniform distribution from 0 to 12 kHz at 90 dB SPL; middle column) is played into the opposite ear. For these trials, a second sound channel was fed to the opposite ear throughout the session and a 90 dB SPL white noise was delivered on each trial beginning 50 ms prior to the initial LED appearing and lasting until 500 ms after the target LED was extinguished. No trials with a click stimulus were used in this control. The contralateral noise reduced the RMS amplitude of the OSEARs in all subjects, although it remained statistically significant in 4 out of the 7 subjects tested (ANOVA, $p<0.05$).

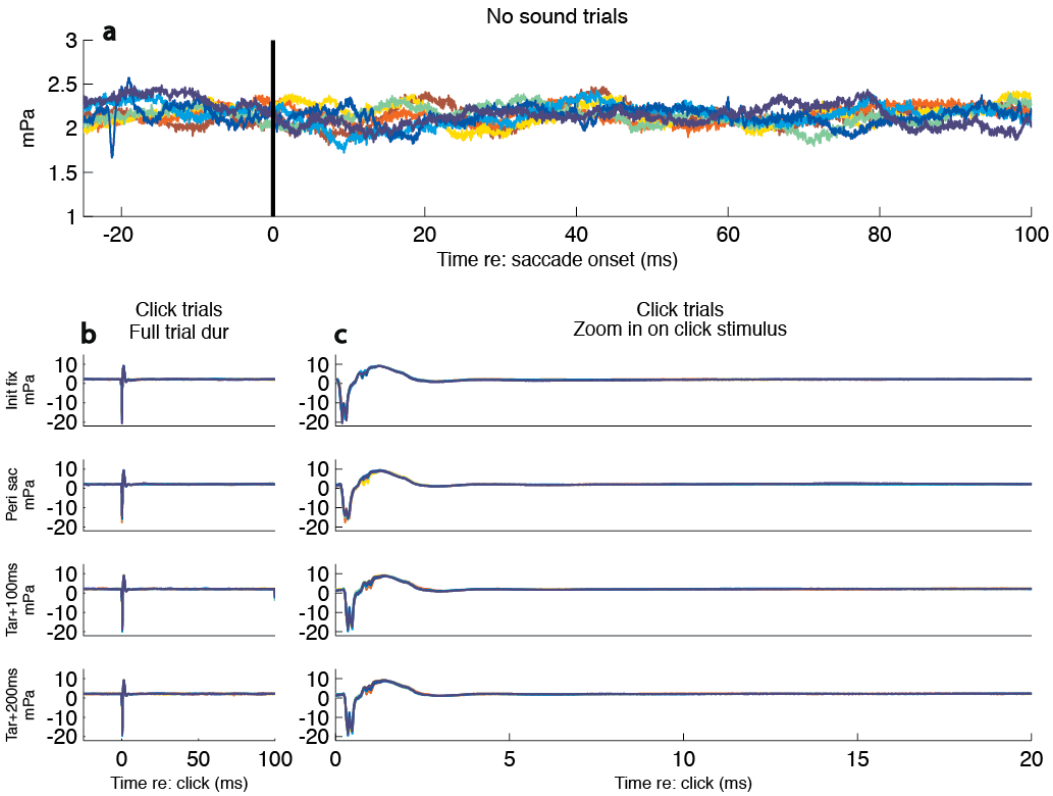

Figure 12: Syringe recording control.

Recordings were made by having a human (HME045) run a full data collection session as per normal, but instead of collecting data from the ear canal, the microphone was set into a syringe with approximately the same volume as a standard ear canal (1mL). The data were otherwise collected exactly as any other session.

2.1.4 OSEARs and click-evoked otoacoustic emissions

We next considered what impact a sound stimulus in the same ear canal has on this process. During half of our trials for human subjects, clicks (65 dB SPL, 40 μs) capable of evoking otoacoustic emissions (CEOAEs; Kemp, 1978) were delivered during the initial fixation, during the saccade to the target, and at both 100 ms and 200 ms after obtaining target fixation. Clicks presented during the initial fixation (figure 13a, top row; inset: zoom in on post-click detail) provide a control case and showed no differences as a function of the upcoming saccade target location, unknown to the subjects at the time of click delivery. This was confirmed using the same point-by-point regression analysis and Monte Carlo simulation as in the no-click trials (figure 13a, third row). Clicks delivered after the eye movement was complete (figure 13c-d, top row), revealed no new sound-triggered effects attributable to the different static eye fixation positions achieved by the subjects by that point of the trial (figure 13c-d, third row).

When clicks were delivered during the eye movement (figure 13b, top row), the post-click microphone signal continued to vary with saccade target location, as it did in silence (figure 13b inset: zoom in on post-click detail; also figure 13b, third row regression analysis). To assess whether this reflected only the basic OSEAR observed earlier, with click and OAE superimposed, or whether the eye-movement related

eardrum motion was changed by the click or its OAE, we eliminated the oscillation by high-pass filtering our data (figure 13a-d, second row; 1kHz cut off; see methods for details) then compared the filtered waveforms. The filtering eliminated the differences previously attributed to the observed OSEARs (figure 13b, second row; compare with 2.12b, top row). We were, however, unable to consistently extract click-evoked OAEs from our recordings from individual mean traces, and not at all at the population level (figure 13, second row, insets). It should be noted that more reliable, multiple click techniques for eliciting OAEs (Probst, 1990) were not feasible to use during a rapid movement of the eyes.

Finally, we tested the peak amplitude of the click itself to see whether there was a difference in the acoustic impedance of the ear canal as a function of saccade direction. Specifically, we were interested in seeing if there were any structural changes in the canal itself that might result from increased thickness of the soft tissue lining the canal wall (e.g. due to contraction of the facial muscles), blood flow at the surface of the canal, or some other similar effect. A reduction in click amplitude would indicate that there was a change within the ear canal such that the canal absorbed more sound than in a more "neutral" state. We found no evidence of this in the peak click amplitude (ANOVA, $p>0.9$; figure 13a-d, bottom row).

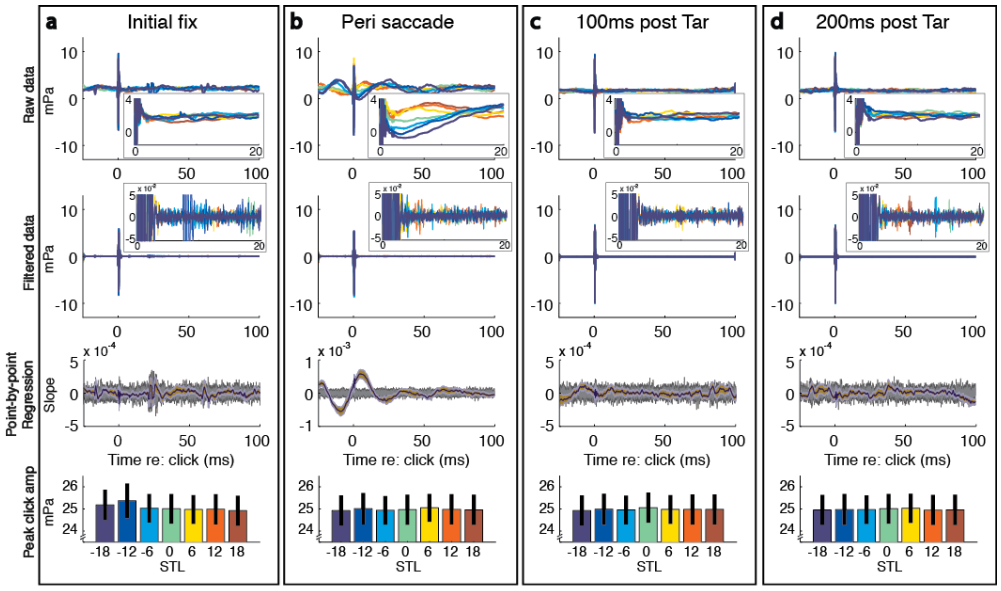

Figure 13: Clicks delivered during but not after eye movements were associated with OSEARs.

Population mean data for human subjects (n=19 ears from 16 subjects). Monophasic clicks with 40μs duration and 65 dB SPL amplitude were presented on 50% of all trials, interleaved, at all of four time points within these trials: (a) during the initial fixation, (b) during the saccade (approximately 20ms after saccade onset), and both (c) 100 ms and (d) 200 ms after target fixation was obtained; each column represents data from one of the click epochs, separated by saccade target location as in figure 4. The first row provides an overview of the raw data from 20 ms before to 100 ms after click onset with an inset zoom in on the peri-click timing for a more detailed view. In either view, it is readily apparent that there is no difference between the traces during the initial fixation (a) where the eyes are always centered, nor is there an effect of static eccentric eye position when the eyes are fixed on the target (c-d). The similarities in the raw data

traces are confirmed with the same regression analysis and Monte Carlo simulation used previously (a, c-d: third row; gold band: real data; gray band: scrambled data). The only difference between the various traces comes in panel b, during the eye movement, which is presumably fully attributable to the already observed effects of the movement itself (note significant regression analysis in panel b, third row). We filtered out the low frequency (f<375Hz) activity for all epochs (a-d, second row with zoomed inset for detailed view) and we observed no additional influence of the click beyond the impact of the eye movement (compare first and second rows, panel b). There were no differences in the peak click amplitudes for any epochs (a-d, fourth row), indicating that the acoustic impedance did not change as a function of saccade target location.

2.2 Discussion

Here, we have demonstrated a regular and predictable pattern of oscillatory movements of the eardrum associated with movements of the eyes, such that when the eyes move left, both eardrums initially move right then oscillate antiphasic from each other for 3 to 4 cycles. This oscillatory saccadic eardrum associated response (OSEAR) is present in both human and non-human primates, is similar across both ears within a subject, and is attenuated by the contralateral ear reflex. Moreover, it is both strong and consistent for individual subjects and can be detected with even as few as 200 trials across 7 saccade target locations (as in the case of monkey subject MHH003).

2.2.1 Why does the eardrum move when the eyes move?

A critical problem in multisensory integration is that sensory systems often provide spatial information in different reference frames and the brain must align these reference frames to form a unified percept of space across the different senses. Determining whether a sight and a sound arise from a common object requires knowing the relationship between the visual eye-centered and auditory head-centered reference frames (Groh & Pai, 2010; Groh & Sparks, 1992) – a relationship that changes whenever the eyes move. Eye position acts as a conversion factor from eye- to head-centered coordinates (K. K. Porter & Groh, 2006), and the auditory system is known to be sensitive to the eyes' position (Bulkin & Groh, 2012a; IC: Groh et al., 2001; superior colliculus: Jay & Sparks, 1984; Lee & Groh, 2012; visual intraparietal sulcus: Mullette-Gillman et al., 2005; Mullette-Gillman et al., 2009; Populin et al., 2004; K. K. Porter et al., 2006; auditory cortex: Werner-Reiss et al., 2003; Zwiers et al., 2004)and movements (IC: Bulkin & Groh, 2012b; LIP: Mullette-Gillman et al., 2009; K. K. Porter et al., 2007). This allows the brain encode auditory space in multiple reference frames (Mullette-Gillman et al., 2005) so that auditory and visual space can be understood within the same reference frame, particularly at the time of an oculomotor command to some visual target (Lee & Groh, 2012). With this study we have demonstrated that the auditory periphery is systematically modulated by eye movements, intimately connecting the

oculomotor system to the entirety of the auditory system. We suspect that this eye movement information helps the auditory system track the eyes' position in order to synchronize auditory and visual space.

At present, it remains a mystery how OSEARs might be used by the system, or what affect, if any, they have on incoming sounds. It is possible that there is no desirable purpose to the OSEAR and that the eye position related signals found elsewhere in the system are independent of this peripheral activity. However, the effect is both strong and consistent, so it seems unlikely that the system could simply "ignore" it, particularly if the eardrum movement actually alters the passage of sound into the auditory system. Instead, it seems that the effect is either 1) an epiphenomenon which the system must actively counteract; 2) an assay of some other neural process that is influenced by eye movements and affects the auditory periphery; 3) an effect whose primary purpose is something other than tracking eye movements but is still closely connected to the movements; or 4) a mechanism that directly supports tracking eye movements in the auditory system. In any case, the effect either directly or indirectly contains information about the eyes' movements that are regular and predictable.

2.2.2 Mechanism(s) responsible for effect

We have shown that eye movements influence the auditory periphery by generating oscillatory movements of the eardrum. Those movements might be best summarized: when the eyes move left, the eardrums initially move right. There are multiple mechanisms within the ear that could potentially be responsible for this finding; in particular, both the OHCs and the MEMs are known to produce or influence acoustic activity in the ear canal. However, OSEARs have a frequency centered near 20 Hz and a peak amplitude of about 3 mPa (or 44 dB) above baseline at the population level, with individual subjects peaking near 15 mPa (57 dB). Sounds generated by OHC activity are typically observed at 500Hz or above, and it is highly improbable that these cells could produce sounds of this magnitude at these low frequencies (Kemp, Ryan, & Bray, 1990).

Our current findings are instead more consistent with MEM activity. While the OSEAR amplitude and frequency profile is inconsistent with previous OHCs recordings, both measures have been well documented with similar ranges in the MEMs. Oscillatory frequencies of 20-40 Hz have been associated with both the tensor tympani and stapedius muscle activity (Borg, 1972a, 1972b; Borg & Moller, 1968; Eliasson & Gisselsson, 1955; Greisen & Neergaard, 1975) and individual muscle fiber contraction

rates for both muscles are also within this range (Teig, 1972). MEM contractions can also generate sound pressure consistent with our recordings in the presence of contralateral sound which occur in the absence of ipsilateral sound (Yonovitz & Harris, 1976). Our recordings show that the eardrum can initially move either inward or outward depending on whether the eyes move toward the contra- or ipsilateral hemifield: when the eyes move left, the eardrums moves right. Importantly, the mechanical structure of the ossicular chain makes it possible for MEM contractions to initiate eardrum movement in either inward or outward directions (Casselbrant, Ingelstedt, & Ivarsson, 1977; Holst, Ingelstedt, & Ortegren, 1963; Yonovitz & Harris, 1976). Furthermore, MEM activity is known to precede self-generated activity such as vocalizations (Avan, Loth, Menguy, & Teyssou, 1992; Carmel & Starr, 1963; Salomon & Starr, 1963). Our results indicate that the observed eardrum movement begins around 10 ms before the eyes actually move, suggesting that an efference copy of the oculomotor command is sent to the mechanisms involved prior to eye movement. All of our data are consistent with all of these previous observations of MEM activity.

Although the MEMs are often associated with attenuating loud environmental and self-generated sounds, they are also known to be active in the absence of explicit auditory stimuli. Specifically, MEM activity during rapid eye movement (REM) sleep is a well-documented occurrence, and tends to produce multi-phasic oscillations that may

occur with or without associated eye movements (Dewson et al., 1965; Pessah & Roffwarg, 1972). It is worth noting that this MEM activity is well correlated, but not in total synchrony, with the eye movements that occur during REM sleep (De Gennaro & Ferrara, 2000; De Gennaro et al., 2000). Additionally, startle reflex activation of the MEMs can be achieved by both cutaneous stimulation of the auditory meatus and by delivery of burst of air to the orbit of the eye (Greisen & Neergaard, 1975; Holst et al., 1963; Salomon & Starr, 1963; Yonovitz & Harris, 1976). Finally, similar to OHCs, the response of MEMs seems to be dependent on behavioral state (Greisen & Neergaard, 1975), and can be influenced by electrical stimulation of the reticular formation (Hugelin, Dumont, & Paillas, 1960). Therefore, the movements we observe are both inconsistent with known OHC physiology and consistent with observed MEM physiology, leading us to believe that MEMs are the more likely candidate mechanism between these two.

It is important to note, however, that we cannot rule out any contribution of the OHCs in this or possible similar effects. OHCs are known to be subject to efferent control (Guinan, 2010) and to modulate peripheral activity as a function of task demands (Delano et al., 2007; Srinivasan et al., 2014; Srinivasan et al., 2012). Specifically, attending an auditory stimulus tends to enhance OHC output while attending elsewhere – for instance to another auditory cue, frequency band, while attending to a visual stimulus – inhibits OHC function (de Boer & Thornton, 2007; Delano et al., 2007; D. W. Smith &

Keil, 2015; P. F. Smith, 2012; Srinivasan et al., 2014; Srinivasan et al., 2012). This similar pattern of activity is held through IC (Rinne et al., 2008) and auditory cortex (Molloy et al., 2015). While we presume that attending to a target is necessary to make an accurate saccade to that target, the clicks in our task were task-irrelevant. Therefore, it is possible that our task may have actually inhibited the click-evoked OHCs activity and that explicitly linking the click to the eye movements in a task-relevant manner may produce different results.

It should also be noted that there is a possibility that the OSEARs are not due to movement of the eardrum at all, but rather to changes in the facial musculature that in turn alter cause the walls of the ear canal to contract or expand and thereby change the volume of the canal. We argue that this is unlikely for multiple reasons. 1) Any change in the soft tissue on the ear canal walls would have to be driven in an oscillatory manner and in synchronized but opposite phase across the two ears by the facial musculature. It seems unlikely that the comparatively large muscles that might be responsible for this could oscillate in this manner, as opposed to undergoing a tonic contraction, and do so in time with each other on opposite sides of the face. 2) There is no obvious reason to believe that the contralateral noise would inhibit the function of any facial muscles that might drive a contraction of the ear canal, while this effect is known to have an inhibitory effect on both the MEMs and OHCs. Therefore, either of these latter two

active mechanisms is a more parsimonious explanation. 3) An alteration of ear canal volume would likely impact the acoustic impedance within the canal. Since we saw no evidence of a change in acoustic impedance at the delivery of the click stimulus, this seems unlikely. 4) The deeper ear canal has rigid, bony walls, and our custom ear plugs were generally set fairly deep into the ear canal. In most cases, this will have set the microphone beyond the fleshier shallow ear canal which would be more susceptible to influences from the facial muscles.

For these reasons, we believe that OSEARs are most likely attributable to the MEMs and that, while OHCs may still be sensitive to eye movements and position, they do not play a significant role in the effects we currently observe.

2.2.3 Concluding remarks

With this study, we have demonstrated that multisensory interactions occur at the earliest possible point in the auditory system, that this interaction is both systematic and substantial, and that the effect occurs in both human and non-human primates. This raises the intriguing possibility that efferent pathways in other sensory systems – for instance, those leading to the retina (Honrubia & Elliott, 1968, 1970; Itaya & Itaya, 1985) – also carry multisensory information to help refine peripheral processing. Importantly,

our data show that the brain integrates early and often in order to make the best informed decision about the world with which it has to interact.

2.3 Methods

2.3.1 Human Subjects and Experimental paradigm

Human subjects (n=16, 8 females, aged 18-35 years; participants included university students as well as young adults from the local community) were involved in this study. All procedures involving human subjects were approved by the Duke University Institutional Review Board. Informed consent was obtained prior to testing, and all subjects received monetary compensation for participation. Stimulus (visual and auditory) presentation, data collection, and offline analysis were run on custom software utilizing multiple interfaces (behavioral interface and visual stimulus presentation: Beethoven software [sampling rate: 500Hz], Ryklin Inc.; auditory stimulus presentation and data acquisition: Tucker Davis Technologies [sampling rate: 25 kHz]; data storage and analysis: Matlab, Mathworks).

Subjects were seated in a dark, sound attenuating room. Head movements were minimized using a chin rest, and eye movements were tracked with an infrared camera

(EyeLink 1000 Plus). Subjects performed a simple saccade task (figure 3a). The subject initiated each trial by obtaining fixation on an LED located at 0º in azimuth and elevation and about 2 meters away. After 200ms of fixation, the central LED was extinguished and a target LED located 6º above the horizontal meridian and ranging from -24º to 24º in 6º intervals was illuminated. The most eccentric targets (±24º) were included despite being slightly beyond the range at which head movements normally accompany eye movements, which is typically 20º (Freedman, Stanford, & Sparks, 1996; Stahl, 2001)) because we found in preliminary testing that including these locations improved performance for the next most eccentric locations, i.e. the ±18 º targets. However, because these (±24º locations were difficult for subjects to perform well, we presented them on only 4.5% of the trials (in comparison to 13% for the other locations) and we excluded them from analysis. After the subject made a saccade to the target LED, he or she maintained fixation on it (9º window diameter) for 250 ms until the end of the trial. If fixation was dropped, i.e. if the eyes traveled outside of the 9º window, at any point throughout the initial or target fixation periods, the data for that trial were excluded from analysis.

On half of the trials, task-irrelevant sounds were presented via the earphones of an earphone/microphone assembly (Etymotic 10B microphone with ER 1 headphone driver) placed in the ear canal and held in position through a custom molded ear plug

(Radians Inc.). These sounds consisted of brief clicks (40 μs positive monophasic pulse) at 65 dB peak-equivalent SPL, and were presented at four time points within each trial: during the initial fixation period (100ms after obtaining fixation, "FIX"); during the saccade (approximately 20 ms after initiating an eye movement, "SAC"); 100 ms after obtaining fixation on the target ("TAR1"); and 200 ms after obtaining fixation on the target ("TAR2").

Acoustic signals from the ear canal were recorded via the in-ear microphone throughout all trials, and were recorded from one ear in 13 subjects (left/right counterbalanced) and from both ears in separate sessions in the other 3 subjects, for a total of n=19 ears tested. Testing for each subject ear was conducted in two sessions over two consecutive days or within the same day but separated by a break of at least one hour in between sessions. Each session involved about 600 trials and lasted a total of about 30 minutes. The sound delivery and microphone system was calibrated at the beginning of every session using a custom script (Matlab) and again after every block of 200 trials. The calibration routine played a nominal 80 dB SPL sound – a click, a broad band burst, and a series of frequencies ranging from 1 to 12 kHz in 22 steps – into the ear and recorded the resultant sound pressure. It then calculated the difference between the requested and produced sound pressures and calculated a gain adjustment profile for all sounds tested. Auditory recording levels were set with a custom software calibration

routine (Matlab) at the beginning of each data collection block (200 trials). All conditions were randomly interleaved.

2.3.2 Monkey Subjects and Experimental paradigm

All procedures conformed to the guidelines of the National Institutes of Health (NIH Pub. No. 86-23, Revised 1985) and were approved by the Institutional Animal Care and Use Committee of Duke University. Monkey subjects (n=3, all female) underwent aseptic surgical procedures under general anesthesia to implant a head post holder to restrain the head and a scleral search coil (Riverbend Eye Tracking System) to track eye movements (Judge, Richmond, & Chu, 1980; Robinson, 1963). After recovery with suitable analgesics and veterinary care, monkeys were trained in the saccade task described above for the human subjects. The trial structure was similar to that used in humans but with the following differences (figure 3b, red traces): (1) eye tracking was done with a scleral eye coil; (2) task-irrelevant sounds were presented on all trials, but only one click was presented at 200-270 ms (jittered range) after saccade target acquisition (slightly later than the TAR2 timing for human subjects); (3) the ±24º targets were presented in equal proportion to the other target locations, but were similarly excluded from analysis as above; (4) initial fixation was 110-160 ms (jittered range), while target fixation duration was 310-430 ms; (5) Monkeys received a fluid reward for

correct trial performance; (6) disposable plastic ear buds containing the earphone/microphone assembly as above were placed in the ear canal for each session (n=2 monkeys tested with both ears in separate sessions, n=1 monkey tested with one ear, for a total of n=5 ears tested); (7) auditory recording levels were set at the beginning of each data collection session (the ear bud was not removed during a session, and therefore no recalibration occurred within a session); (8) monkeys' sessions were not divided into blocks (the monkeys typically performed consistently throughout an entire session and dividing it into blocks was unnecessary); and (9) the number of trials per session was different for monkeys versus humans (which were always presented exactly 1200 trials over the course of the entire study); the actual number of trials performed varied based on monkey's performance tolerance and capabilities for the day. Monkeys MNN012 and MYY002 were both run for 4 sessions per ear (no more than one session per day) over the course of 2 weeks; MNN012 correctly performed an average of 1071 out of 1212 trials per session for both ears, while MYY002 correctly performed 788 out of 1305 trials per session on average. Monkey MHH003 was only recorded one day for 200 correct trials (25 trials per STL); upon visual inspection but before analysis, these data were deemed to be consistent with other subject data and were therefore included for analysis. It is worth highlighting that the effect reported in this paper can be seen, in this case, with only a few trials.

2.3.3 Control sessions

To verify that the apparent effects of eye movements on ear-generated sounds were genuinely acoustic in nature and did not reflect electrical contamination from sources such as myogenic potentials of the extraocular muscles or the changing orientation of the electrical dipole of the eye ball, we ran a series of additional control studies. Some subjects (n=4 for plugged microphone control; n=7 for contralateral noise control; n=1 for syringe control) were invited back as participants in one or more of these control studies (3 subjects participated in both the plugged microphone and contralateral noise studies).

In the first of these control studies, the microphone was placed in the ear canal and subjects performed the task but the microphone was physically plugged, preventing it from detecting acoustic signals (see figure 11a). This was accomplished by placing the microphone in the custom ear mold as usual, but the canal-side opening for the microphone was blocked. Thus, the microphone was in physically the same position during these sessions, and should therefore have continued to be affected by any electrical artifacts that might be present, but its acoustic input was greatly attenuated by the plug. 4 subject ears were re-tested in this paradigm in a separate pair of data collection sessions from their initial "normal" sessions. The control sessions were

handled exactly as the normal sessions except that a second, plugged ear mold was used to replace the first, open mold after the microphone was calibrated in the open ear. Calibration was executed exactly as in the normal sessions prior to plugging the ear mold.

The second control experiment once again used the same general trial structure but the session was modified. The primary difference in this versus the normal paradigm was that a 90dB SPL white noise burst (uniform distribution, 0 to 12kHz) was played into the contralateral ear for each trial beginning 50 ms before the initial fixation onset and lasting through the entire duration of the trial, through 500 ms post-target fixation offset. The noise level was set according to the calibration for the opposite (recorded) ear. Each of the two sessions per subject in this study consisted of 3 blocks of only 100 trials presented with no less than 10 minutes of rest between block; no trials with click stimuli were included in these sessions. We included the extended rest period and reduced the number of trials by eliminating the click-stimulus trials in order to allow the active mechanisms in the ear some time to recover between blocks and avoid fatigue. Additionally, the loud sound was fairly uncomfortable for all subjects and reducing the number of trials made this more tolerable.

A final control study was run in which the microphone was set into a 1mL syringe, which is the approximate average volume of the human ear canal. In this experiment, the trials were run exactly as a normal session except that the microphone was set into the syringe using a plastic ear bud, and the syringe was place on top of the subject's ear behind the pinna. Acoustic recordings were taken from within the syringe while a human subject executed the behavioral paradigm exactly as normal. Sound levels were calibrated to the syringe at the start of each block.

Unless specifically stated otherwise, all data are reported using the raw acoustic sound pressure recorded from the microphone (converted from volts to Pascals after recording). Results for individual human and monkey subjects were based on all of that subject's correct and included trials (with ±24º target locations excluded, the average number of correct trials per human subject was 932 ±104 standard deviation; average number of trials per target location = 161±18).

We used two criteria to identify and discard especially noisy trials. A trial was rejected if (1) that individual trial had a standard deviation that exceeded 10 times the standard deviation of the entire block; or (2) any individual sample within that trial had a magnitude greater than 50 times the session standard deviation. In either case, these trials were typically due to movement or bodily noise (e.g. swallowing, chewing, and so

forth). These criteria only eliminated a small proportion of trials: 7.1±5.7 correct trials per subject (mean ± standard deviation).

Saccades were identified in each trial using a custom routine (Matlab). The routine identified any region in the raw eye trace where the acceleration of the eye was greater than $7^{o}/sec^{2}$, then refined that search by isolating movements where the velocity was greater than 55^{o}/sec. The identified saccades were then aligned to the acoustic data by using a series of time points triggered from one source (Beethoven presentation control routine) and marked separately in the acoustic and eye position data. Any trials were there were potential communication errors identified by misalignment of the state markers were discarded. This check rejected 5.5±2.6 trials per subject (mean ± standard deviation). The eye position data were multiplied by -1 for all subjects whose left ears were recorded in order to convert eye movement data from left/right coordinates into contra/ipsilateral coordinates. Finally, each trial's saccade target location (STL) was verified by comparing the LED target presented by Beethoven with the mean target fixation location and confirming that these two numbers matched.

2.3.4 Statistical analyses

We measured the significance of each individual subject's data by comparing the sound pressure associated with each STL for all of that subject's trials. We used the first prominent peak of the population mean waveform (at t=5.5 ms; see figure 3b) and, for each trial for a given subject, calculated the mean sound pressure from t=4.5 to t=6.5 ms. We compared these mean values with an ANOVA, where STL was the grouping factor for the ANOVA, for all of that subject's trials.

We ran a Monte Carlo simulation for each subject to determine the effect size relative to baseline error rate for all statistical tests used on these data. In this routine, all trials at sample j, where $j = \{1, \ldots, J = 3052\}$, were analyzed with linear regression with sound pressure p_j was the dependent factor and STL for each trial was the grouping factor. Each test produced a set of values – p-value v_j, R^2 r_j, and slope s_j – for each of the 3052 tests run across the data set. This produced a time-series vector of scores for each measure for each subject.

Because we are running a series of statistical tests, and we do not know the dependence of one sample p_j relative to the next sample p_{j+1}, a post-hoc correction (e.g. Bonferroni correction) was not practical. Instead, we used a Monte Carlo technique to estimate the chance-related effect size and false positive rates of this test. We first

scrambled the relationship between each trial and its STL assignment, then ran the same analysis as before. This provided an estimate of how our results should look if there was no relationship between STL and the acoustic recordings.

In order to compare the amplitude between hemifields, we calculated the root mean squared (RMS) amplitude per STL for each subject using the mean data trace per STL for each subject. We then took the mean value for all STLs < 0 for the contralateral hemifield and all STLs > 0 for the ipsilateral hemifield and compared these values for the population (n=19) using a t-test.

For the within hemifield comparison, we calculated the RMS values as before, but in this case we also calculated the true mean saccade end point per STL for each subject (the mean value of the eyes after target fixation has been obtained for each LED target per subject) in order to capture any within subject idiosyncrasies. We then ran a regression on the population RMS values against the absolute value of the mean saccade end points.

Click-trial data were analyzed initially on the raw microphone sound pressure using the point-by-point regression analysis and Monte Carlo simulation previously described. These trials were then high-pass filtered (6^{th} order Chebyshev filter, 375 Hz

cut off [80 dB attenuation] with half-octave roll off) and inspected for click-evoked otoacoustic emissions both at the individual mean and population mean levels.

Peak click amplitudes were calculated for each trial by isolating the maximum and minimum peaks during the click stimulus of the high-pass filtered data. The amplitudes of both scores were averaged to obtain the power of the click stimulus. This allowed us to inspect for a change in acoustic impedance in the ear canal.

3. Anticipation of an upcoming sound alters acoustic output of the ear

3.1 Introduction

That sights and sounds interact in the brain has been known for some time, and these interactions are known to impact multiple facets of auditory perception. Interactions such as sound localization (and humans: Alais & Burr, 2004; in owls, e.g.: Bergan & Knudsen, 2009; Bertelson, Frissen, Vroomen, & de Gelder, 2006; Brainard & Knudsen, 1993; Frissen, Vroomen, & de Gelder, 2012; Gutfreund et al., 2002; Jack & Thurlow, 1973; Kopco et al., 2009; Recanzone, 1998; Thurlow & Jack, 1973; Thurlow & Rosenthal, 1976; Vroomen, Bertelson, & de Gelder, 2001a, 2001b; Vroomen, de Gelder, & Vroomen, 2004; Vroomen & Keetels, 2006, 2009; Woods & Recanzone, 2004), speech recognition (e.g. Bertelson, Vroomen, & De Gelder, 2003; Campbell, 2008; MacDonald & McGurk, 1978b; McGurk & MacDonald, 1976; Pare, Richler, ten Hove, & Munhall, 2003; Soto-Faraco & Alsius, 2009), and multimodal attention (L. Busse, Roberts, Crist, Weissman, & Woldorff, 2005; Laura Busse & Woldorff, 2003; Donohue, Green, & Woldorff, 2015; Donohue, Roberts, Grent-'t-Jong, & Woldorff, 2011; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010; Weissman, Warner, & Woldorff, 2004) are of great importance and interest, particularly with the rise of competing environmental stimuli

like text messaging and portable music devices. Traditionally, these sorts of studies tended to focus on interactions at the level of association cortex (e.g. Chow & Hutt, 1953), a trend that continues today (Ghazanfar & Schroeder, 2006). However, more recent work has also focused on trying to establish where and how the sensory systems begin to interact with emphasis on finding the earliest point, meaning most peripheral, point of convergence between sensory systems.

The presence of descending control mechanisms from the brain to the auditory periphery make it conceivable that the visual system can actually modulate the mechanics of the complex of physical structures associated with auditory transduction. There are various routes through which visual information could pass to the auditory periphery, mostly by way of the inferior colliculus (IC). The IC acts as an important hub in the auditory system through which nearly all ascending and descending fibers in the system pass (Winer & Schreiner, 2005). This region is influenced by numerous non-auditory processes, including vision (Bulkin & Groh, 2012b; Gutfreund et al., 2002; Mascetti & Strozzi, 1988; K. K. Porter et al., 2007). Many of the efferent pathways that route into through the IC originate in brain regions – notably the retina (Herbin et al., 1994; Itaya & Van Hoesen, 1982; Yamauchi & Yamadori, 1982; A. B. Zhang, 1984), visual cortex (Cooper & Young, 1976), and superior colliculus (Adams, 1980; Coleman & Clerici, 1987; Covey et al., 1987; Hyde & Knudsen, 2000; Stitt et al., 2015) – that are

important for vision (for review, see Gruters & Groh, 2012). Once in the auditory system, these visual signals could conceivably be sent to various mechanisms in the periphery that are involved in the transduction of auditory information.

Another clue pointing towards interactions between the visual and auditory systems affecting the auditory periphery is our recent discovery that the eardrum vibrates when the eyes move (see Chapter II of this document). In this study, humans and monkeys made eye movements to the locations of visual targets. A period of eardrum oscillation accompanied the eye movements (Oscillatory Saccadic Eardrum Associated Response, or OSEAR). These OSEARs were tightly locked to the time of occurrence of the saccade, and they varied systematically in phase and magnitude depending on the direction and amplitude of the saccade. The tight relationship between OSEARs and the spatial properties of the eye movements suggested that they play a role in generating a common spatial reference frame between the auditory head-centered and visual eye-centered reference frames. Accordingly, we sought to determine if visual stimuli exert an influence on peripheral auditory processing that is distinct from this eye-movement related peripheral signal.

Within the auditory periphery, there are multiple active mechanisms that may respond to audiovisual interactions. Of particular interest are the outer hair cells (OHCs)

of the cochlea, and the muscles of the middle ear (MEMs). Both the OHCs (Guinan, 2006, 2010) and MEMs (Mukerji et al., 2010)receive substantial efferent input and are sensitive to non-auditory influences. MEMs are known to contract in response to startling acoustic and cutaneous stimuli as well as self-generated sounds (Carmel & Starr, 1963; Djupesland, 1964; Greisen & Neergaard, 1975; Salomon & Starr, 1963; Yonovitz & Harris, 1976)}; they are also known to contract during REM sleep (De Gennaro & Ferrara, 2000; De Gennaro et al., 2000; Dewson et al., 1965; Pessah & Roffwarg, 1972). Meanwhile, the response amplitude of OHCs is inhibited at the population level when subjects attend visual stimuli as compared to auditory stimuli (de Boer & Thornton, 2007; Delano et al., 2007; Ferber-Viart et al., 1995; Froehlich et al., 1993; Meric & Collet, 1992, 1994; Meric et al., 1996; Puel et al., 1988; D. W. Smith et al., 2012; D. W. Smith & Keil, 2015; Srinivasan et al., 2014; Srinivasan et al., 2012). However, it is not clear if simply engaging in a visual task globally inhibits these cells' response to incoming sounds or if the effect is more nuanced, and the involvement of the MEMs in this process, if any, is unknown.

Both the OHCs and MEMs are capable of indirectly generating movement in the tympanic membrane, so ear canal sound pressure may be used as an indirect measure of both OHC and MEM activity. We took advantage of this phenomenon and recorded sound pressure within the ear with a microphone placed in the canal to investigate whether visual stimuli that are predictive of upcoming sounds exert an influence over

the state of the auditory periphery. We designed a behaviorally simple paradigm in order to minimize the inhibitory attentional effects associated with the cognitive load of a complex visual task. Additionally, we were interested in changes to the "baseline" state of the system as opposed to how the system responds to a sound stimulus; accordingly, we also designed the paradigm to allow us to compare two trial types that had no sound stimulus. To do this, we compared trials in which a sound was expected with those where no sound was expected. This allowed us to determine if the visual component of the audiovisual stimulus could explicitly cue the auditory component regardless of the actual presence of that component.

3.2 Methods

Human subjects (n=15, 8 females, aged 18-35 years; participants included university students as well as young adults from the local community) were involved in this study. All procedures involving human subjects were approved by the Duke University Institutional Review Board. Informed consent was obtained prior to testing, and all subjects received monetary compensation for participation.

3.2.1 Set up and visual display

Testing for each subject took place over two sessions run one each on two consecutive days. Each session consisted of 400 trials split into 100 trial blocks and took approximately 40 minutes of testing. Subjects were given the opportunity to leave the testing booth between blocks if necessary in order to minimize fatigue throughout testing. Each subject was fit with a custom molded ear plug (Radians Inc.) prior to the first day of testing; that same plug was used for all subsequent testing done by that subject. A combination earphone/microphone assembly (Etymotic 10B microphone with ER 1 headphone driver) was placed into the ear plug to deliver auditory stimuli and record acoustic pressure in the ear canal. Sounds were controlled by a high speed processor (Tucker Davis Technologies, RX6), and recorded sounds were stored offline for analysis. Sound levels were calibrated at the start of each session (or block, if the subject removed the ear plug between blocks) using a custom routine (Matlab). The calibration routine played a nominal 80 dB SPL sound into the ear and recorded the resultant sound pressure. It then calculated the difference between the input and output amplitudes and calculated a gain adjustment profile for all sounds tested. Sounds were calibrated for clicks, broad band burst, and a series of frequencies ranging from 1 to 12 kHz in 22 steps.

During a session, the subject was seated with a chin rest in a dark, sound attenuating room approximately 65 cm away from a standard LCD computer monitor. Eye movements were monitored during each trial with a tripod mounted video eye tracker (Eye Link 1000), which was calibrated at the start of each session; calibration was adjusted in the Eye Link software manually using multiple fixation points on the computer screen set to known positions. To initiate each trial, subjects fixated on a target in the center of the screen (0°, 0°), and they were required to maintain fixation near the center of the screen throughout the trial (within a 4.5° radius window).

Each trial consisted of a short video (1.5s duration; 60 frames at 40 fps; videos were generated and displayed using custom software [Matlab]) that started with two discs, colored in different and distinguishable hues of gray, spaced approximately 9º visual angle away from each other horizontally and 5.5º above the fixation point.

In all trials, the discs moved downward across the screen over the course of the trial to end at 9° horizontal spacing and 5.5° below the fixation point. The downward trajectory was either straight down ("drop" type trials) or angled to where the discs approached each other for the first half of the trial and made contact at the fixation point, then traveled away from each other for the second half of the trial ("bounce" type

trials). Note that there were no trials in which the discs streamed through each other, as the different hues of the discs made it clear that they had "bounced off" each other.

Three different trial types were presented. Audiovisual (AV) trials (figure 14a) occurred on 70% of trials (trial types randomly interleaved). During these trials, the discs moved downward and inward across the screen for the first half of the trial, made contact at the fixation point, and then traveled away from each other for the second half of the trial to end at 9° horizontal spacing and 5.5° below the fixation point. A brief sound played when the two discs made contact at the fixation point (75 dB SPL, open source sound emulating two colliding billiards balls). Visual-only bounce-type (vB) trials (figure 14b) occurred on 15% of all trials and were the same as AV trials except that no sound was played at the visual contact point. The remaining 15% of trials were visual-only drop-type (vD) trials, wherein the discs moved in a straight downward trajectory and ended in the same location as both AV and vB trials; no sound was played. Regardless of whether or not an auditory stimulus was played, sounds were recorded from within the ear canal for the duration of every trial (figure 1d shows the mean population data traces for AV [black], vB [blue], and vD [red] trials). At the end of each trial, subjects reported whether the trial was a "standard" (AV) or "deviant" (vB or vD). They were asked to withhold their response until the entire video was complete. After that, they had 1s to register their response with a keypad.

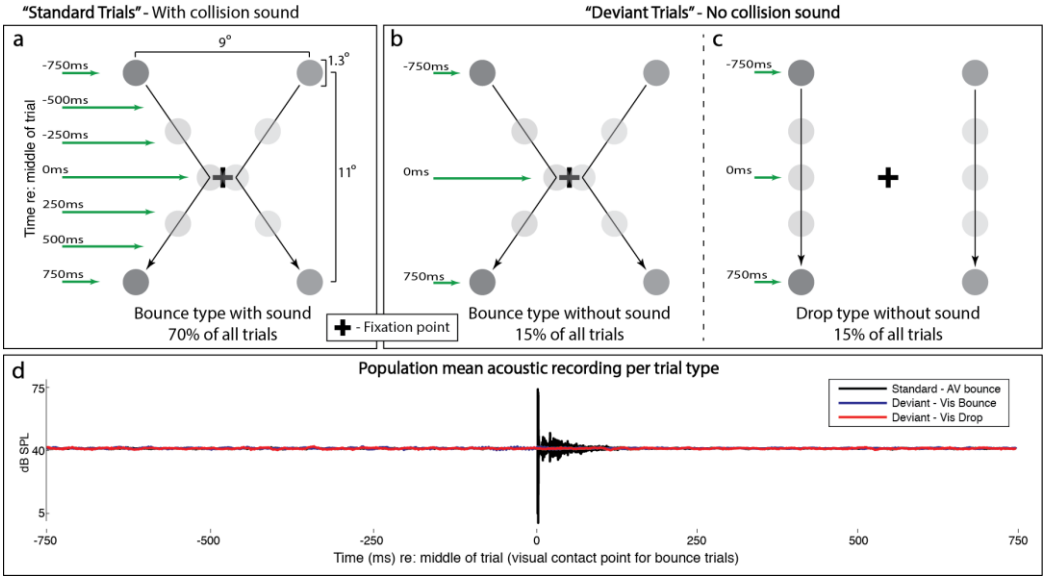

Figure 14: Trial schematic.

All trials began with two disks (approximately 1.3º diameter) located approximately 9º away from each other horizontally and 5.5º above the central fixation point (0º,0º); disks were colored distinguishable hues of gray. Three possible trial types were used in this experiment. In standard (a) trials, the disks moved downward across the screen while they approached each other, made contact at the fixation point, then separated; a collision sound like two billiard balls contacting each other (75 dB SPL; black trace in panel d shows population mean acoustic data trace for standard trials) was played when the two disks touched. These trials occurred for 70% of all trials and were intended to initiate an expectation of sound when the disks moved on a colliding trajectory. There were two deviant type trials which each occurred on 15% of all trails (b-c). Visual only bounce (vB) type trials (b) used the same animation as the standard trials but had no contact sound (blue trace on panel d; note the lack of explicit auditory stimulus). In visual

only drop (vD) type trials (c), the disks moved downward across the screen but did not approach each other; there was no contact sound for these trials (red trace in panel d). Each trial took 1.5s (60 frames at 40fps), with the disks ending at 5.5º below fixation. All timings are relative to the contact point in the bounce type trials, or the same relative time in the drop type trials (referred to as "T0" for the duration of this report).

A small group of subjects (n=3) was invited back to participate in a follow up experiment. This study used the same trial structure, but used a tone pip (generated in Matlab; 50ms duration, 10ms ramp on and off; 75 dB SPL) as the auditory stimulus instead of the billiard ball contact sound. 400 trials were conducted each of 5 consecutive days, where each day a different tone frequency – 0.5, 1, 2, 4, or 8 kHz – was used. Tone presentation order was randomized across subjects, but each subject only heard one tone each day. Otherwise, the paradigm presentation and execution was the same.

3.2.2 Analysis

For each subject, trials were included for analysis if they were: 1) answered correctly (AV trials identified as "standard", as well as vB and vD trials identified as "deviant"); 2) answered when prompted but not before (after the completion of the full trial); and 3) fixation was maintained within a 4.5º radius window around the fixation

point for the duration of the trial (for visual only trials, mean number included is 244±11 standard deviation; for AV trials, mean number included is 536±12 standard deviation; there were no differences between the number of included vB and vD trials; performance was always 90%+ correct). Analysis timing is all relative to the onset of the center-most frame of the video for each trial, denoted "T0" for time = 0ms. For the bounce trials, the center time was the point of visual contact between the balls, while in the drop trials, the duration of the trial was identical, so the center time point occurred as the ball was exactly half way along its path; in the auditory trials, this was the point where the sound was played.

Where applicable, trials' frequency spectra were computed using a fast Fourier transform (FFT) routine (Matlab) across either the entire trial or in 200 ms epochs spanning from -700 to +700 relative to T0 (i.e. -700 to -500 ms, -500 to -300 ms, and so forth). Data in these analyses were expressed as a difference in decibel levels between trial types, with the decibels being calculated against the pre-trial average sound pressure for all visual trials included for a given subject. Specifically, the dB value x_j was defined for each sample j as $x_j = 20log_{10}(\frac{p_j}{R})$, where p_j is the sound pressure recorded by the microphone for that sample, and R is the reference value defined as the root mean squared amplitude for the full trial duration from all trials included in that subject's

analysis. This normalization was applied to allow comparison across subjects with different background "noise" levels in the ear canal.

We ran a Monte Carlo simulation on the population level data to determine the effect size relative to baseline error rate for our frequency-series (FFT) based data. We compared the mean frequency power for all subjects at each sample f_j, where $j = \{1,2,\ldots,J\}$ and J = the number of samples in our frequency series, in our FFT series with a paired samples t-test such that, for each test, n=15 each for vB and vD type trials. This resulted in a series of p-values associated with each frequency sample 1 through J in our data set.

Because we are running a series of statistical tests, and we do not know the dependence of one sample relative to the next sample, a post-hoc correction (e.g. Bonferroni correction) was not practical. Instead, we used a Monte Carlo technique to estimate the chance-related effect size and false positive rates of this test. We scrambled the trial type assignment for each of the mean data traces in our dataset (30 total, 15 each vB and vD) then ran the same analysis as before. We repeated this scramble procedure 10 times to determine the expected range of error that would be produced for a random set of trial type assignments. This provided an estimate of how our results should look if there was no relationship between the trial type and the acoustic recordings.

3.2.3 Estimation of entropy through generalized variance

We measured the difference in generalized variance, $V(x)$, for each visual trial type by calculating the determinant of the covariance matrix, $det[Cov(x)]$, separately for vB trials and vD trials per subject. This measure of variance provides a single digit value for the complete multidimensional scatter within a dataset, and is an optimal estimator of signal entropy (Cai, Liang, & Zhou, 2013). We were interested in this measure as a function of the frequency ranges identified by our FFT analysis, however, so we first filtered (6^{th} order Chebyshev filter with quarter octave roll-off) the data into three frequency bands: low-pass (f < 100 Hz), band-pass (100 Hz < f < 4 kHz), and high-pass (4 kHz < f), as well as unfiltered data. Next, we generated a separate $n \times n$ covariance matrix for each subject's vB data and vD data, where n = number of trials, using the raw data (microphone amplitude in Pascals, no normalization). As per the definition of a covariance matrix, each cell in this matrix represents the covariance of two trials, where row 1 is the covariance of trials {1,1}, {1,2}, … {1,n}; row two is the covariance of trials {2,1}, {2,2}, … {2,n}; and so forth. Because the autocovariance of a signal is equal to its variance, the diagonal of this matrix is the variance of trials 1 through n, and the matrix is symmetrical across the diagonal. We then subtracted the natural log of the square root of the determinant of vD trials from vB trials (equation 1).

In order to avoid having to scale either matrix, we limited the number of trials in the larger of the vB or vD datasets to the number of trials in the smaller set by randomly rejecting the difference in number of trials from the larger data set. For instance, if a subject had 105 vB trials and 100 vD trials, we randomly rejected 5 vB trials. This resulted in two separate covariance matrices – *Cov(vB)* and *Cov(vD)* – that were both the same *n x n* size. This ensured that there were no differences in the generalized variance measure due to simply having more trials while also minimizing potential rounding errors that resulted when applying a scaling factor in the calculation of the determinant. This eliminated 15 trials on average out of an average total of 122 trials with no tendency to eliminate more from one dataset than the other across the population. This same routine was done for each 200 ms time epoch as well as for the full trial duration.

Equation 1

$$V(vB) - V(vD) = \frac{1}{2}\ln\left(\frac{\det[Cov(vB)]}{\det[Cov(vD)]}\right)$$

This analysis provides a more precise measure of the variance that is strictly due to trial type. While a standard approach to calculating mean variance includes many "unknown" sources of noise that may increase the variance of the signal – such as movement artifacts or acute environmental sounds – the covariance matrix calculation minimizes these additional sources of variance while preserving the variance introduced by the independent variable.

3.2.4 Statistical testing

Statistical tests in all cases with reported p-values were 2-tail t-tests based on the difference in each subject's mean data for vB – vD against a null hypothesis of no difference. Specifically, for each subject, the mean vD score was subtracted from that same subject's mean vB score, and the population level ttest was calculated for n=15 values. In the case of the FFT data, mean scores for each subject represent the mean difference for a given frequency range. For instance, the score for a given subject in the mid-range FFT results is generated by taking that subject's mean difference for all frequencies 0.1kHz < f < 4kHz.

3.3 Results

It is important to note, at the outset of these comparisons, that we are analyzing the acoustic "noise" in the ear canal for all trial types. There is no explicit sound stimulus; instead, we are recording the combined activity of the various mechanisms in the ear as they subtly change the air pressure within the ear canal.

3.3.1 Comparison of visual trial types' frequency content

We examined the frequency content of the visual only trials by executing a fast Fourier transform (FFT) on each trial for the full trial duration (figure 15a), as well as multiple epochs defined with respect to the point of visual contact during bounce type trials (figure 15c, top row). We then subtracted the mean FFT values for the vD trials from the vB trials for the FFT taken during the designated epoch. With this analysis, we see that these two trial types differ from each other, but their relationship is different depending on which frequency range we observe; there is little difference across the various temporal epochs tested.

We organized our analysis into three frequency ranges based on where the difference score of vB-vD crosses a zero value. We then compared the mean difference score across each frequency range against a null hypothesis of no difference (t-test, n=15). When the FFT values were calculated based on the full trial duration, there were differences in each frequency band ($p<0.05$; figure 15a). Specifically, both the low- and high-frequency ranges had a greater amplitude for vB than vD trials, while mid-range data had a lower amplitude for vB than vD trials. We further assessed the profile of these differences by comparing each data sample in the frequency series with a separate t-test, and compared these results to an "expected error rate" determined through an

iterative Monte Carlo simulation (figure 15b; see methods for details). While our mid- and high-frequency data are consistently below p=0.05 for the bulk of the range in question, the low-frequency data are more sporadic and noisy. The population level differences seen within each epoch are generally well reflected at the level of individual subjects (figure 16), particularly at about 250 Hz and above.

We further examined the mean difference in FFT power within a series of 200 ms time bins (figure 15c). The bar plot (figure 15c, top row) represents the mean difference in FFT power between vB and vD for the population within the frequency ranges noted. Specifically, we averaged the FFT power for each trial type within the specified frequency range and bin for each subject and compared the set of means (t-test, n=15). With this analysis, we see a trend emerge that the effects are not existent in any frequency range at the beginning of the trial, but as the trial progresses, the difference between the two trial types becomes more pronounced. The difference reaches maximum from 100 to 300 ms after T0, where the two balls would make a contact sound in AV trials, then returns toward baseline over the rest of the trial.

This trend might emerge due to changes in one or both trial types. For instance, we may get these results if 1) vD trials increased in mid-range frequency power, 2) if vB trials decreased in mid-frequency power, or 3) some combination of effects. The bottom

row of figure 15c shows the mean power (dB re: pre-trial sound pressure) for both trial types, and we can see that vD trials remain near baseline (start of trial mean for all vB and vD trials) for the duration of the trial in all frequency ranges while vB trials deviate in all ranges. This indicates that the active mechanisms involved in this process are influencing the ear during vB trials but not vD trials.

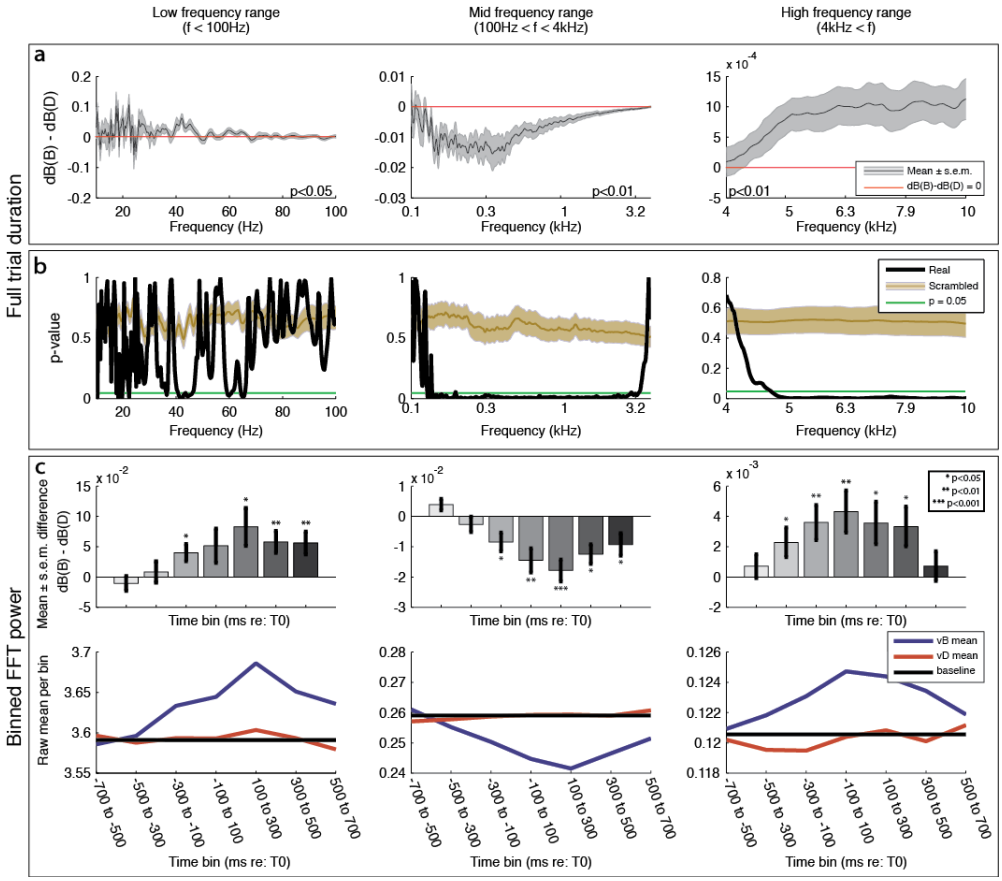

Figure 15: Population level analysis of frequency power for deviant (vB and vD) trial types.

A fast Fourier transform (FFT) was run on each trial across either the full trial duration (a-b), and a series of shorter epochs throughout the trials (c). Differences in amplitude between deviant trial types fell into three distinct frequency zones, shown in separate panels to allow the data to be viewed with magnification of both axes appropriate for each frequency range. Statistical tests in panels a and c (upper row) represent the mean amplitude difference within the entire frequency range shown in each panel. Panel b gives a more precise method of looking at the profile

of differences between trial type data; this test confirms that the data within the prescribed frequency bands are significantly different from around 250 Hz up through 3.2 kHz for the mid-frequency range and above 5 kHz for the high-frequency range. Mean data traces in panel c, lower row, show that the active mechanisms involved in this effect influence vB trials (blue trace), while vD trials (red trace) remain near the start-of-trial baseline (black trace). Low frequency activity (f<100Hz): leftmost column in both panels; vB > vD. Mid-frequency activity (0.1kHz < f < 4kHz): center column; vB < vD. High-frequency activity (4kHz < f): rightmost column; vB > vD again.

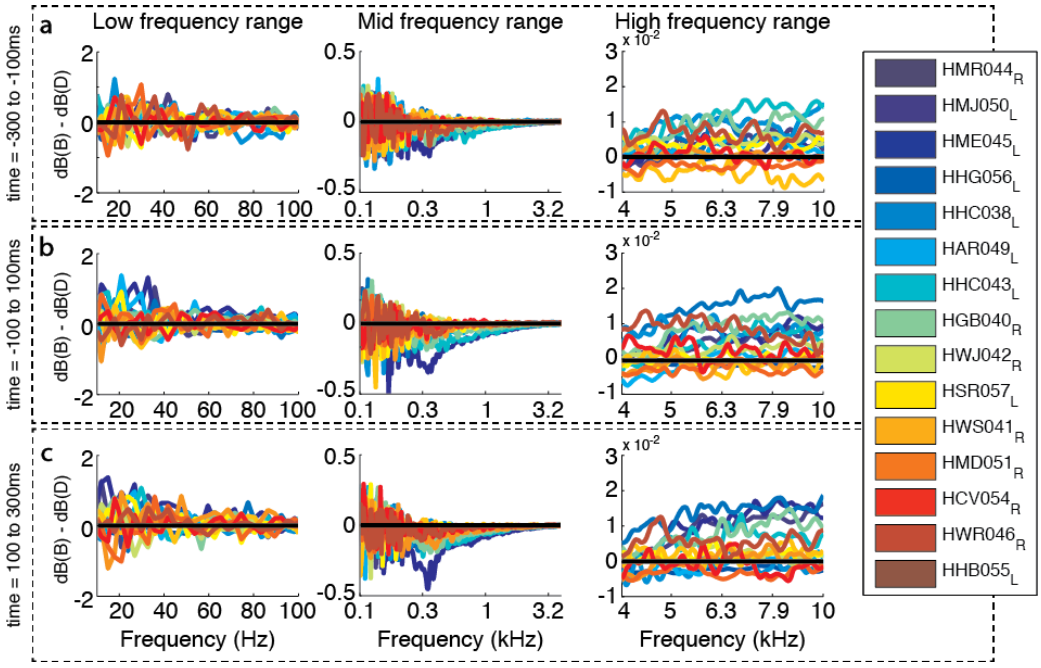

Figure 16: Differences (vB – vD) in individual subject FFT results for three epochs near T0.

Results for each individual were typically consistent with the rest of the population. Traces in the low frequency window (left column) and the low end of the mid-frequency window (middle column) tended to be quite noisy relative to the higher frequency data.

As a reality check, we compared the time period from -300 to -100 ms between the AV and vB trials. These trials are visually identical and there has been no sound yet on the AV trials. The information available to the subjects prior to the sound stimulus is the same, so the statistical differences observed in the vB – vD comparison should be

absent. Figure 17 shows that this is the case: the AV-vB difference is not significantly different from 0.

As a positive control to assess the reproducibility of the findings, we also compared the AV trials to the vD trials during the same -300 to -100 ms silent time period. In this case, the results should replicate the vB-vD comparison during this temporal epoch, and they do. The high frequency AV-vD difference is greater than 0 ($p<0.05$; Figure 17b, left panel) and the mid-frequency AV-vD comparison is less than 0 ($p<0.01$; Figure 17b, right panel). This confirms that the frequency content for both of the bounce type trials (AV and visual only) is similar up until the point of the sound onset in the AV trials.

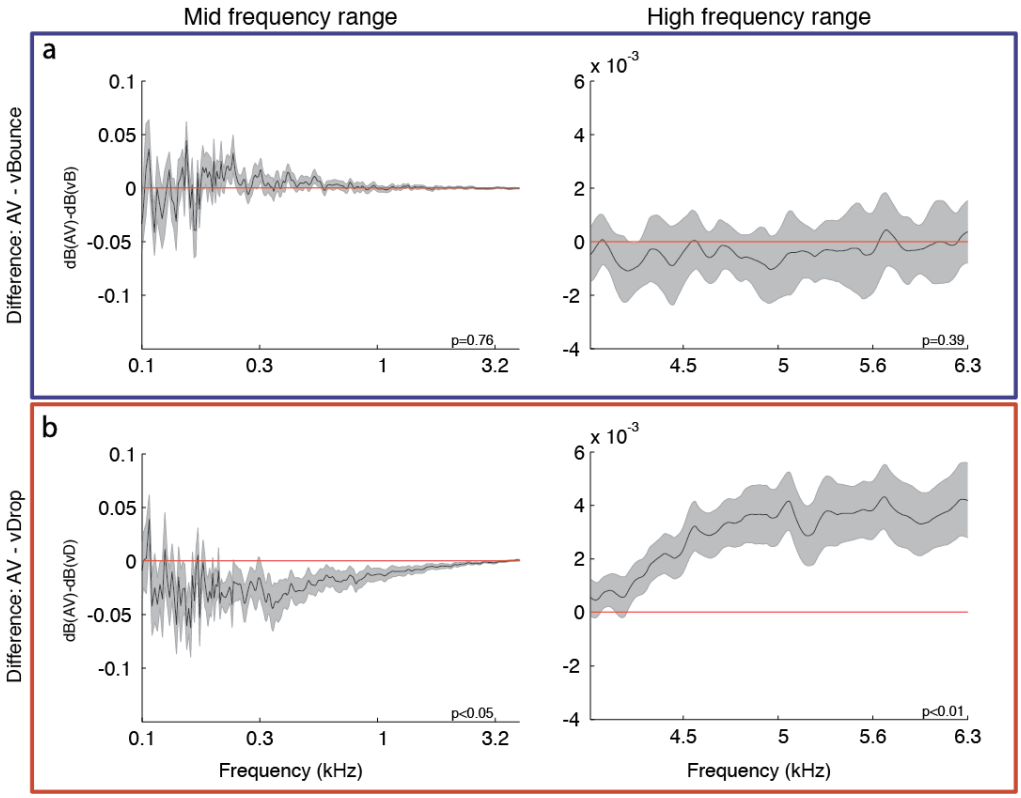

Figure 17: Population mean difference in standard versus deviant frequency power for time period -300 to -100 ms.

The information leading up to the contact point for both the standard audiovisual (AV) and deviant bounce (vB) trials should be the same to the observer because there is no way to distinguish in advance what type of trial it is. Therefore, these two trial types should be, and are, statistically similar (a). Likewise, the difference between the AV trials and deviant drop (vD) trials should be similar to the vB – vD differences seen in figure 15, and this is the case (b).

3.3.2 Comparison of generalized variance

One of the primary roles of the peripheral auditory system appears to be optimizing signal to noise ratio (D. W. Smith & Keil, 2015). Accordingly, active peripheral processes may be expected to actively engage in order to minimize noise and allow a signal to more easily be detected by the system. One measure of the predictability of a system (or, the relative presence or absence of noise) is its variance. The determinant of covariance for a system – or the generalized variance, as an estimator of signal entropy – is a means of quantifying variance across some unknown number of dimensions within a dataset, particularly biological and environmental noise unassociated with the test paradigm. Because we are trying to isolate the variance introduced into our measurements specifically by the trial-type parameter while (nearly) eliminating all other sources of variance, this method is a more accurate and refined measure of variance for our purposes.

We tested the generalized variance, $V(x)$, of the previously defined frequency ranges for our data by applying, separately, a high-pass, band-pass, and low-pass filter to our visual trial data, as well as testing the entire frequency range with unfiltered data. For the full trial duration, we found that the variance of the high-frequency ranges was not statistically different for vB versus vD trials (t-test, $p>0.5$; figure 18b), but for the mid-and low-frequency ranges and unfiltered data, vB trials had significantly less variance than vD trials (t-test, $p<0.01$; figure 18a, c-d). Throughout the series of 200 ms time epochs, the high-frequency data remained insignificant (t-test, $p>0.2$; figure 18f). For the other frequency filters as well as the unfiltered data, there was a trend for vB trials to become progressively more different than vD trials up until 300 ms after T0, then return toward baseline (see p-values noted on figure; figure 18e, g-h).

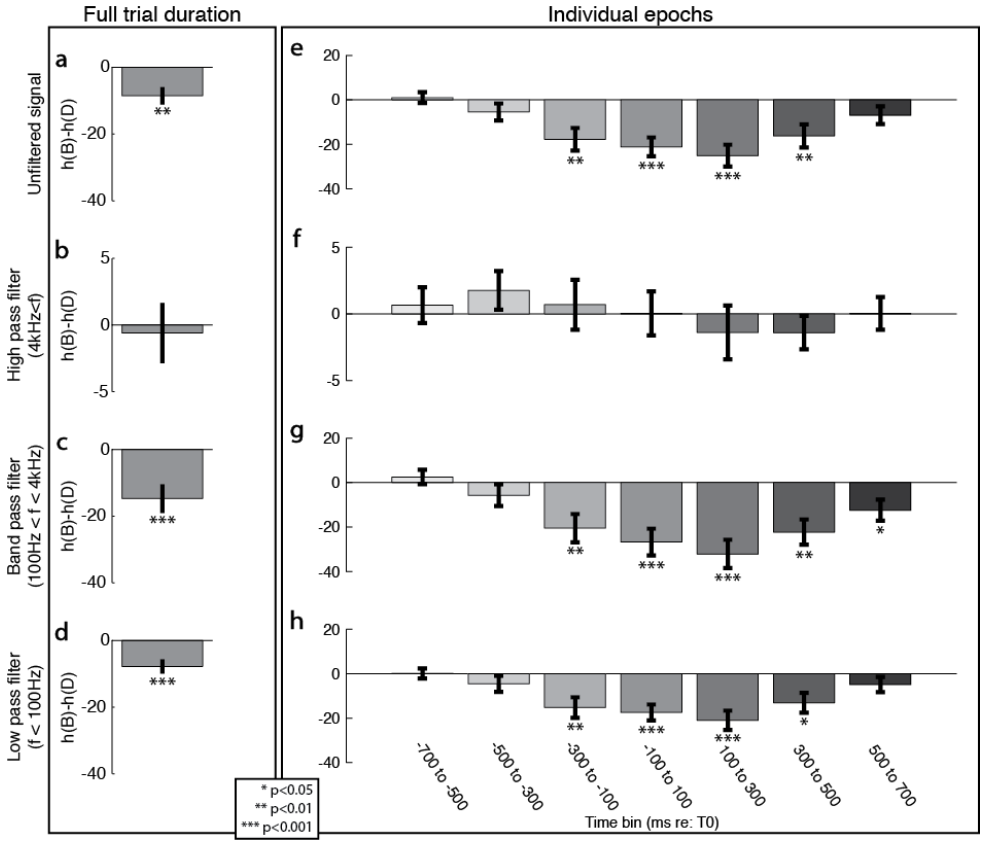

Figure 18: Population mean differences (vB – vD) in generalized variance, V(x).

The determinant of covariance (see methods for details) was calculated for unfiltered (top row) and filtered data (high-pass, second row; band-pass, third row; low-pass, bottom row) in order to measure the difference in generalized variance between both trial types. There were no differences in variance between trial types in the high frequency information for the full trial duration (b), or any individual epoch (f). However, the unfiltered, band-pass filtered, and low-pass filtered data were all significant across the full trial duration (a, c-d). Over the course of the

trial, these frequency ranges all began with similar variance, then became progressively more dissimilar until shortly after T0 before returning toward baseline (e, g-h).

Individual subjects' data in both the low-pass (figure 19, left column) and band-pass (figure 19, middle column) filtered data had consistently lower generalized variance in all three epochs immediately surrounding T0 (i.e., -300 to -100 ms, -100 to 100 ms, and 100 to 300 ms). Notably, only two subjects (HWR046$_R$, and HHB055$_L$) went consistently against the population across all epochs in these lower two frequency ranges, and only one other subject (HMD051$_R$) also showed greater variance for vB trials in any case (both lower frequency bands during early epoch [figure 19a], and low frequency only during peri-T0 epoch [figure 19b]). In the high frequency range, the same three subjects had greater variance for vB than vD trials for each of the shown epochs, in addition to subject HHG056$_L$. Each epoch in this frequency range also had only one other subject with a similar directionality to their difference score, and that subject was different for each epoch. So although the high-pass filtered data is quite insignificant at the population level, that effect is largely driven by a specific subset of the population.

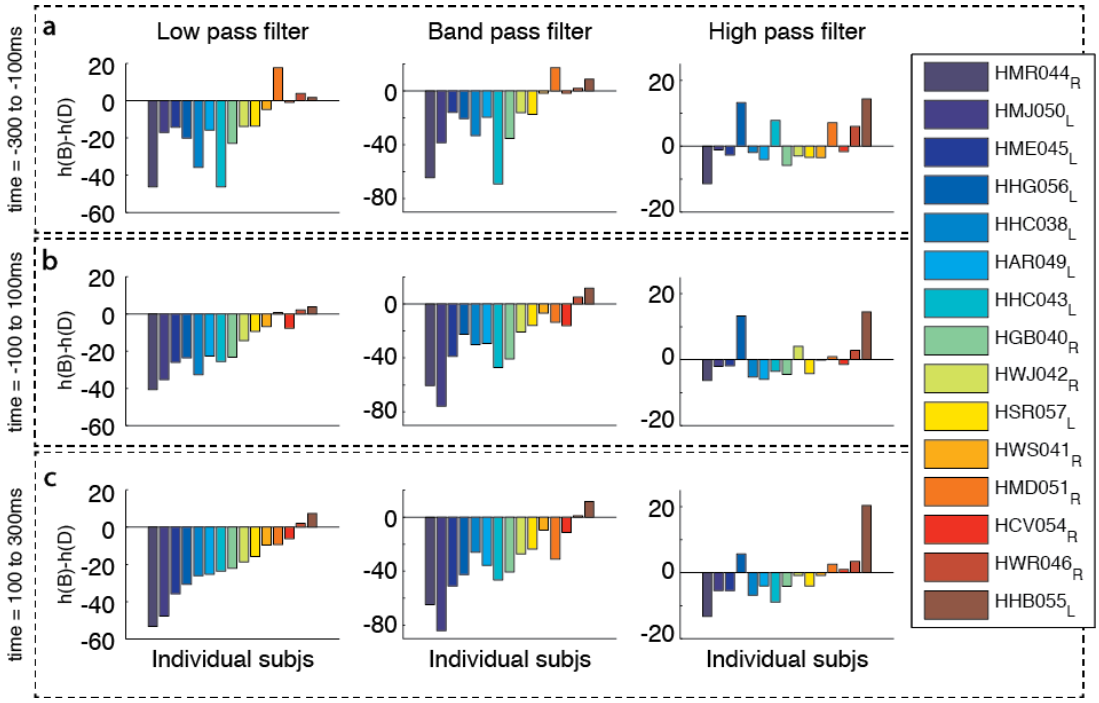

Figure 19: Difference in generalized variance scores (vB – vD) for all individual subjects around T0.

Most subjects had greater variance in vD than vB trials during the three epochs surrounding T0; 8 out of 15 subjects always measured differences in this direction, while 3 of the remaining 7 only showed anomalies in one epoch each of the high-pass filtered data. Subjects HWR046R and HHB055L had greater variance in vB than vD trials consistently, while HMD051R also had a score in this direction for all comparisons in the epoch immediately before T0 (a) and all high-pass comparisons (right column); HHG056L had this effect for all high-passed data as well.

3.3.3 Effects are not frequency specific with stimulus

The sound stimulus we used during the initial paradigm had a broadband spectrogram with frequencies especially concentrated around the 2-3 kHz region. Given that the differences we observed between vB and vD trials varied across the frequency spectrum, we sought to determine if the frequency of the anticipated sound affected the visually-guided preparations of the auditory periphery. Accordingly, we ran 3 of the subjects on a series of additional sessions in which the contact sound's frequency was varied. The contact sounds were tone pips at either 0.5, 1, 2, 4, or 8 kHz and only one frequency was used in a given session. Each subject performed one session per day for five days.

For the mid-frequency data, we found that the frequency of the tone pip could influence the magnitude of the vB-vD differences, but did not shift the frequency-ranges in which those vB-vD differences occurred (figure 20, left panel). However, the tone pip data generally exhibited lower vB amplitude relative to vD trials for both the mid- (left panel) and high- (right panel) frequency data. This pattern was not modulated by the specific frequency of the tone pip in an obviously systematic fashion. Instead, it seems likely that the overall stimulus power, which was stronger and more diffuse across frequency space in the original sound versus the tone pips, may be responsible for this

finding. In particular, there does not appear to be any tendency for the "tuning curve" of the vB-vD difference to shift with contact-sound frequency.

This test suggests that anticipating of a sound of a particular frequency affects some aspects of the visually-cued differences in auditory peripheral activity, but does not suggest that the brain is suppressing activity selectively for the anticipated range of frequencies. More data involving a greater number of subjects and frequencies might reveal more subtle effects.

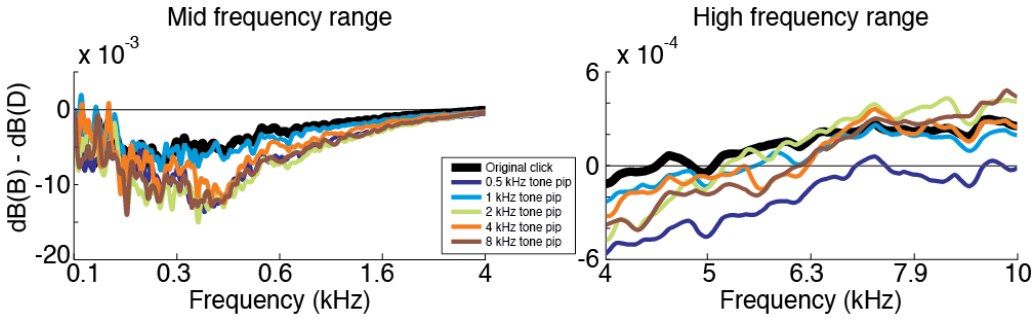

Figure 20: Comparison of mean (n=3) data from tone pip stimulus paradigm variants.

In these trials, the original contact sound (thick black trace) was replaced with a tone pip (50ms, 10ms ramp up and down; 75 dB SPL) at one of 5 frequencies per session with each tone presented over a course of five sessions on consecutive days. Data here represent the difference (vB – vD) in full trial duration FFT power (compare with figure 15a). For the most part, the tone

pip data closely resemble the original click data for this set of subjects. While there are differences in the mean traces, there is no consistent and predictable effect of tone pip frequency. The most notable difference between tone pips and the original stimulus is in the 4 to 6.3kHz range, where tone pips seem to generally have less power than the click. However, there is no ordered gradation within the pips and no tendency for the frequency range in which a difference occurs to shift for different tone pip frequencies.

3.4 Discussion

The purpose of this study was to determine whether a visual stimulus predictive of a task-relevant auditory stimulus can induce changes in the activity of the auditory periphery. We found that trials where a sound is expected have less power at mid-range frequencies (from 100 Hz < f < 4kHz) and increased power at low (f < 100 Hz) and high (4kHz < f) frequencies as compared to trials where a sound is not expected. Moreover, there is less signal variance for frequencies less than 4kHz when a sound is expected, and this difference becomes more pronounced around the time that the sound is expected. This question is important for many real-world audiovisual situations which require rapid and succinct interaction between the auditory and visual systems.

There are many cases in day-to-day life where it may be advantageous to use a visual cue to selectively enhance some associated auditory cue. Detecting speech in

noise – as per the well-known “cocktail party” effect – is one such example. In many of these cases, it would be useful to prevent the passage of background noise into the auditory system while allowing the expected signal to pass through, and vision may be the best cue that such a peripheral filter ought to be applied. That we find evidence in support of this in the noise-floor of the ear canal is indicative of this being a regular and natural process, rather than an uncommon occurrence applied only in select circumstances.

3.4.1 Functional role

Our data are consistent with one of the prevailing theories on function role of the auditory periphery: that the mechanisms involved globally inhibit the auditory system to increase signal to noise ratios (SNR) (D. W. Smith & Keil, 2015). In particular, the reduced variance for vB trials, where a sound is expected, supports the concept of reducing the uncertainty in the auditory system to allow a signal to emerge. Moreover, the continued reduction in variance around the time of the contact point suggests that the system may continue to "search" for a sound around some broad time window wherein that sound might occur. Regardless, SNR reduction is thought to be accomplished by means of descending cognitive control on peripheral mechanisms via the MOC (ibid.).

While a great deal of scientific research has been dedicated to understanding how the brain filters out noise in order to amplify a signal, relatively little work has focused on how it

limits the amount of noise that first enters the system. This is particularly true of various peripheral mechanisms excluding the OHCs, despite there being evidence that these mechanisms are under the influence of multiple non-acoustic processes. Our results indicate that this is likely a much more active process, involving many diverse cognitive and sensory cues as well as multiple active mechanisms in the auditory periphery, than has traditionally been assumed.

3.4.2 Source of signal

Acoustic data recorded from the ear canal represent the combined activity of a number of active mechanisms within the ear; anything that can cause a change in the air pressure of the ear canal serves as a potential source of activity in these recordings. We have intentionally recorded in such a way as to not limit the contributions of one source versus another because there is no a priori reason that any one source should not be involved. Both the OHCs and MEMs are well studied mechanisms in the ear and have previously been linked to non-auditory influences on auditory processing. Other potential mechanisms, including the auricular muscles around the pinnae and annulus fibrosus musculature the eardrum, are not well studied and cannot be easily ruled out.

Additionally, we have recorded in the absence of a sound stimulus in order to explicitly look at changes to the "baseline" acoustics of the system. This affords the opportunity to look at mechanisms more generally. Previous studies that have shown

influences of visual attention on the periphery have used sound stimulus to activate OHC function, which may be only a part of the overall peripheral filter process. Here, we show that there are changes to the system that are apparent even in the absence of sound, that presumably influence the system when a sound is present.

Our data do not strongly argue for or against any specific peripheral mechanism. The OHCs are known to amplify attended frequencies while being inhibited for unattended frequencies (D. W. Smith et al., 2012; Srinivasan et al., 2014; Srinivasan et al., 2012). Given this, we would have expected to see differences in our data associated with the stimulus frequency, but we saw no such effect. Such an effect would have been strong evidence for the involvement of OHCs in this effect, but this lack of evidence does not negate their possible influence on these findings.

The MEMs may also be involved in these present findings. Contraction of the MEMs pulls on the ossicles, which, due to their complex mechanical synergy, exert non-orthogonal force on the tympanic membrane (TM) (De Greef et al., 2014; Huttenbrink, 1988, 1989; van den Berge, van Geest, Rensema, & Drukker, 1990); the dynamics of the ossicular motion suggest that contraction of the MEMs could result in asymmetrical tension being applied to the TM. Because different subdivisions of the TM vibrate in different frequency ranges (Cheng et al., 2010; De Greef et al., 2014; Khaleghi, Furlong,

Ravicz, Cheng, & Rosowski, 2015; Kunimoto et al., 2014; Rosowski, Cheng, Merchant, Harrington, & Furlong, 2011; X. Zhang et al., 2014), it is possible that a contraction of the MEMs alone could explain the frequency dynamics we observe in our data, or that they could work in tandem with the OHCs to produce our findings.

Regardless of whether MEMs, OHCs, or both mechanisms are involved, it seems likely that when the visual system indicates that a sound is expected (e.g. during a Bounce type trial), the mechanisms involved contract slightly in order to minimize the influence of potentially interfering sounds and optimize signal detection. This is consistent with the proposed role of the medial olivary complex (MOC) and its control over the OHCs in similar scenarios.

4. McGurk effect in the auditory periphery: The influence of mismatched audiovisual speech cues on active mechanisms in the ear

4.1 Introduction

Visual cues in speech carry a substantial amount of information for the listener. Lip movements are particularly helpful as they are redundant with auditory speech and begin prior to the associated sound with a latency of 100-300ms (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). This redundancy and lead time allow the brain to use visual cues to help inform the interpretation of the auditory cue, particularly when that cue is degraded due to background noise or hearing impairment (Duchnowski, Hunke, Busching, Meier, & Waibel, 1995; Sumby & Pollack, 1954).

The McGurk effect (McGurk & MacDonald, 1976) is a well-known illusion that takes advantage of this interaction by presenting a video of a speaker visually saying some syllable while dubbing over the audio with a different, but similar, speech sound. The visual cue typically partially "captures" the auditory cue, forcing the auditory percept to sound somewhere in between the visual and auditory cues. Two commonly used examples are the Ba-Ga and Ba-Fa pairings; in both cases, the sound /ba/ is paired with a video of a speaker saying either "Ga" or "Fa". The Ba-Ga pairing most often

elicits a percept of /da/ (MacDonald & McGurk, 1978a), while an /f/ percept in Ba-Fa pairings strongly overrides the /b/ sound (Yonovitz, Lozar, Thompson, Ferrell, & Ross, 1977).

Processing of audiovisual speech, including the McGurk effect, is known to involve numerous cortical regions in both humans (for review, see: Campbell, 2008) and primates (Romanski & Diehl, 2011; Sugihara, Diltz, Averbeck, & Romanski, 2006). For instance, (Beauchamp, Nath, & Pasalar, 2010) demonstrated that application of transcranial magnetic stimulation can disrupt McGurk perception, and (Skipper, van Wassenhove, Nusbaum, & Small, 2007) showed that the cortical activation for fused percepts more closely matches the motor plan associated with the fused syllable than the constituent syllables. However, the auditory system has a substantial descending pathway that is capable of carrying visual information to brain regions earlier in the system (Winer & Schreiner, 2005). The inferior colliculus, which is known to have numerous non-auditory influences and is sensitive to vocalization-related processes (Aitkin, Tran, & Syka, 1994; Champoux et al., 2007; Champoux et al., 2006; Fischer, Bognar, Turjman, & Lapras, 1995; Klug et al., 2002; Pincherli Castellanos, Aitoubah, Molotchnikoff, Lepore, & Guillemot, 2007; Suta, Kvasnak, Popelar, & Syka, 2003; Tammer, Ehrenreich, & Jurgens, 2004), may act as a conduit for speech information to pass all the way out to the auditory periphery. Our lab has recently shown that the

eardrum oscillates in conjunction with movements of the eye (see Chapter II of this document) and visual cues appear to prepare the auditory periphery for an upcoming sound (Chapter III). Other work has shown that the outer hair cells of the cochlea modulate cochlear sensitivity in a frequency selective manner (e.g. Srinivasan et al., 2014); it is therefore conceivable that visual speech cues may prime the system to receive the auditory cues that follows.

We studied whether visual speech cues in McGurk stimuli alter the activity of the auditory periphery in response to an acoustic speech stimulus. Different visual videos were dubbed with the exact same audio speech sounds, which were played directly into the ear canal while a microphone simultaneously recording the sound pressure within the canal. These recordings captured the aggregate of the sound stimulus and the acoustics of the ear canal, which is modified by the brain's various descending auditory control mechanisms. Comparison of the recordings for the same sound paired with different videos isolated the contribution of changing ear canal acoustics and revealed that the recorded sounds differed in accordance with the illusory speech percepts elicited by the different videos.

4.2 Methods

Human subjects (n=22, 10 females, aged 18-35 years; participants included university students as well as young adults from the local community) were involved in this study. All procedures were approved by the Duke University Institutional Review Board. Informed consent was obtained prior to testing, and all subjects received monetary compensation for participation.

4.2.1 Set up and task structure

Subjects participated in a single session consisting of 600 trials split into 75 trial blocks; each session lasted for approximately 90 minutes of testing. Subjects were given the opportunity to leave the testing booth between blocks if necessary in order to minimize fatigue throughout testing. At the start of the session, each subject was fit with a custom molded ear plug (Radians Inc.). A combination earphone/microphone assembly (Etymotic 10B microphone with ER 1 headphone driver) was placed into the earplug to deliver auditory stimuli and record acoustic pressure in the ear canal. Sounds were controlled by a high-speed processor (Tucker Davis Technologies, RX6), and recorded sounds were stored offline for analysis. Sound levels were calibrated at the start of each session (or block, if the subject removed the ear plug between blocks) using

a custom routine (Matlab). The calibration routine played a nominal 80 dB SPL sound into the ear and recorded the resultant sound pressure. It then calculated the difference between the input and output amplitudes and calculated a gain adjustment profile for all sounds tested. Sounds were calibrated for clicks, broadband burst, and a series of frequencies ranging from 1 to 12 kHz in 22 steps.

During a session, subjects were seated with a chin rest in a dark, sound attenuating room approximately 65 cm away from a standard LCD computer monitor. Eye movements were monitored during each trial with a tripod mounted video eye tracker (Eye Link 1000), which was calibrated at the start of each session; calibration was adjusted in the Eye Link software manually using multiple fixation points on the computer screen set to known positions. To initiate each trial, subjects fixated on a target in the center of the screen (0°, 0°), which disappeared during the trial, and they were required to maintain fixation near the center of the screen throughout the trial. After acquiring fixation, on 90% of the trials a short video of a speaker (male; age 29) saying "Ba Ba", "Ga Ga", or "Fa Fa" was played (figure 21a; 2s video duration; 60 frames at 30 fps; videos were generated and displayed using custom software [Matlab]). On the remaining 10% of the trials, a still image from the "Ba Ba" video, taken just prior to the articulation of the /b/ sound, was played for the same duration (figure 21a, yellow framed image). All conditions were randomly interleaved. Each video or still image was

accompanied by the same auditory component: the syllable /ba/ repeated twice (figure 21b; 75dB SPL, ~1000ms audio duration; video started 430ms before audio). The same sound file was dubbed over the accompanying visual component for all videos, including the "Ba Ba" video, as well as the still image.

The video or still image was presented in a square frame, approximately 13º visual angle height and width. Subjects were required to maintain fixation within a 5.5° radius window around the original fixation location; this window was large enough to allow eye movements to the speaker's mouth and eyes, both of which are commonly fixated during natural speech (Buchan, Pare, & Munhall, 2007). At the end of the video or still image, subjects were asked to report the sound that most closely matched what they perceived in a three-alternative forced choice task with the options /ba/, /da/, or /va/. All subjects were specifically instructed to report what they heard, and were told that what they heard might not perfectly match the video or choice options. They were asked to withhold their response until the entire video was complete. After that, they had ~1s to register their response with a keypad.

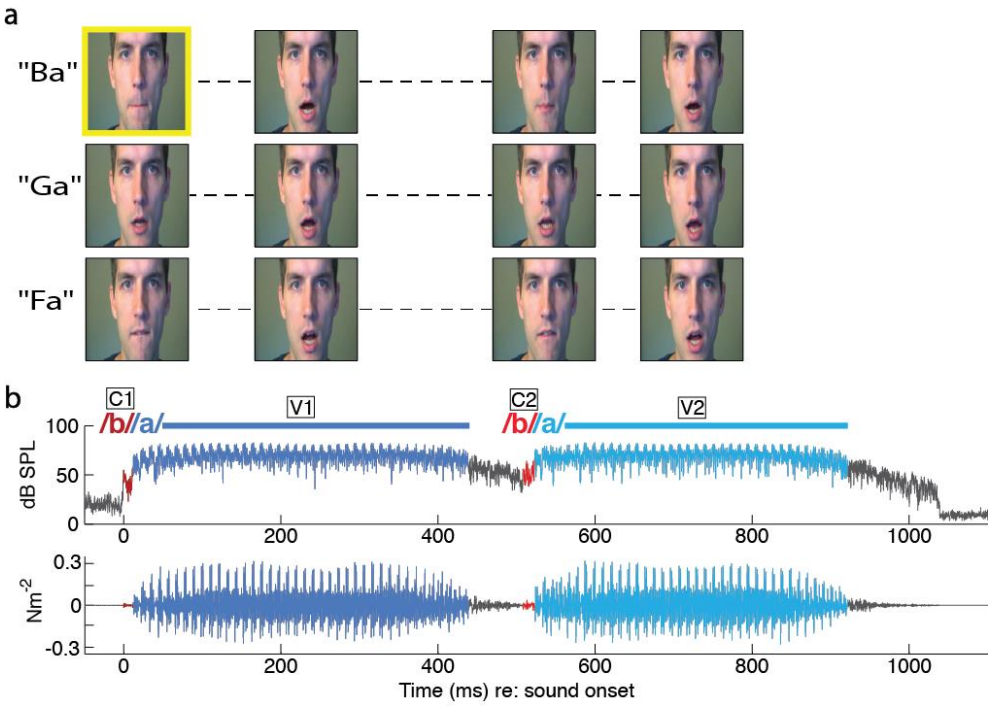

Figure 21: Stimulus details and analysis zones.

(a) The visual component of each trial was a 2-second long video (60 frames at 30 fps; square frame, approximately 13º visual angle height and width) of a speaker saying one of three syllables "Ba", "Ga", or "Fa", twice per trial; these syllables each occurred with a frequency of 30% of trials. Frames from each video at the onset of the consonant articulations and mid-way through the steady state of the vowels are shown here. The other 10% of trials consisted of a static image (yellow framed image in the "Ba" video) displayed for 2 seconds; these trials were included as catch trials and were not analyzed.

(b) The auditory component of each video was an identical recording of a speaker saying /ba/ (75 dB SPL, ~1000ms duration beginning 430ms after video onset). This sound was recorded

separately from the videos and dubbed into each, including the "Ba" videos. All analyses in this paper are focused on the consonant sounds (b; red colored regions), as these are the areas that distinguish the actual speech sounds Ba, Ga, and Fa from each other. Note that the timing in this figure is relative to sound onset; the videos extended for approximately 500ms before and after speech onset and offset with no sound stimulus. At the end of each trial, subjects were asked to pick which sound most closely resembled what they heard given the three alternative forced choices of "Ba", "Da", and "Va".

4.2.2 Analysis

Trial types throughout this document are referred to by their visual components, hence: Ba, Ga, and Fa type trials.

All analyses were run using only trials where subjects 1) maintained fixation in the defined 5.5º radius window; 2) indicated a fused audiovisual percept, i.e. a response of /ba/ to "Ba" videos, /da/ to "Ga" videos, and /va/ to "Fa" videos (mean 165±8 standard deviation trials per video per subject); and 3) delivered answers when prompted after the completion of the video (i.e. trials were not analyzed when subjects answered too early).

Analysis regions of interest were defined based on the raw sound stimulus. This sound clip was inspected and manually annotated prior to running any analyses (figure 21b, colored regions). Consonants (red annotations) and vowels (blue annotations) durations were defined separately with respect to both syllables, where C1 is the first consonant, C2 the second, and V1-2 are the two vowels in sequence; consonant periods were 12 and 15ms respectively, consistent with times previously recorded for the voiced bilabial plosive /b/ (Edwards, 1981), while vowels were 409 and 382ms. These timings were subsequently used to define all analysis windows. Pre- and post-stimulus baseline periods were defined using a 15ms window (specifically for comparison with the consonant analyses) starting 45ms before sound onset and 45ms after the cessation of the second syllable respectively.

In order to minimize variance between subjects and recording blocks for all analyzes, we first normalized raw sound pressure data (Pascals) to zscores within a recording block using only the trials from that block that were included in the final analysis (75 trials ≥ number of trials). To do so, z_j was defined for each sample j as the air pressure p_j recorded by the microphone minus the mean pressure for all analyzed trials per block, μ, divided by the standard deviation for the same trials, σ: $z_j = \frac{p_j - \mu}{\sigma}$. We then isolated the consonants, vowels, and baseline periods for analysis.

For each analysis, we calculated the values of interest for each consonant, vowel, and baseline period. We then averaged the two scores (C1 and C2, V1 and V2, or Pre-base and Post-base) for each trial within a trial type (Ba, Ga, and Fa), and then averaged across all trials within a trial type for each subject. Finally, because we are interested in the degree of difference between the recordings associate with congruent Ba videos versus the incongruent Ga or Fa videos, we subtracted each subject's mean Ba score for a given analysis from their mean Ga and mean Fa scores. Therefore, all analyses are calculated based on population (n=22) mean values.

Because our McGurk stimuli differed according to their elicited *consonant* percept, we focused our analyses on these consonants. In particular, we are interested in comparing the elicited percepts, which have the same "ba" auditory component and differ only in their visual components, with the "true" spectral values of the perceived speech sounds. Previous literature has found that when a voiced bilabial plosive (/b/) is paired with a visual fricative ("F"), subjects perceive a fricative pairings (Yonovitz et al., 1977); typically, an auditory /ba/ paired with visual "Ga" results in a perception of the voiced alveolar plosive /da/ (MacDonald & McGurk, 1978b). Phonetically, both /b/ and /d/ are plosive consonants, which are characterized by a stoppage of sound during articulation and therefore have generally lower sound power at all frequencies, have low amplitude, and low variance during the consonant articulation. In contrast, the fricative

consonants have relatively strong sound power (van Son & Pols, 1999), with fairly high amplitude and variance (Maniwa, Jongman, & Wade, 2009). Finally, the phoneme /v/ also allows the fundamental voiced frequency to sustain throughout the articulation of the consonant, such that the sound power in the frequency range of 100-300Hz should be stronger than that of a plosive consonant despite all three consonants being voiced. Accordingly, our analyses have been structured to inspect these differences across the percepts elicited by the three videos. Specifically, we would expect the Fa-Ba comparisons to be relatively different while Ga-Ba comparisons should be fairly similar, reflecting the "true" differences between the perceived phonemes /va/, /da/, and /ba/.

We used two frequency-based analyses for our data: Fast Fourier transform (FFT) and coherence. Briefly, coherence is a measure of how similar two signals are at all frequencies present in either signal. More specifically, it measures the spectral similarity of two signals by comparing their crossed- and auto-spectral densities (Stoica & Moses, 2005). The crossed-spectral density is simply an FFT of the cross-correlation function between two signals, while the auto-spectral density is an FFT of the auto-correlation function of one signal and itself. These spectral density measures therefore give the distribution of power shared by these two signals at all frequencies present in both signals. Coherence is the ratio of the square of the cross-spectral density to the product of the two auto-spectral densities.

For the coherence measure, we tested all trial types (individually per trial) against the raw auditory stimulus file. This gave the amount of energy preserved by the recording from the (intended) auditory stimulus input to the recorded sound output. Coherence will be affected by the absorption of sound energy by the ear, injection of sound energy into the signal by active mechanisms within the ear, and the transfer function of the electronic hardware (i.e. the microphone and speaker set up). A perfect coherence score of 1 would be obtained in the case of no signal loss in electronics and perfect reflection of the stimulus from speaker to microphone recording.

We tested the statistical difference of both FFT and coherence values between trial types by averaging the Fa – Ba and Ga – Ba difference scores across the frequency range of interest. Specifically, we looked at the frequency range wherein most important speech frequency information lies, at 0.1 to 4kHz; additionally for FFT analyses, we measure the typical approximate pitch range of the human speech fundamental, typically around 100-300 Hz (Hillenbrand, Getty, Clark, & Wheeler, 1995).

In order to better understand the profile of the frequency-based analyses, we also ran a Monte Carlo simulation on the population level data to determine the effect size relative to baseline error rate for our frequency based data (FFT and coherence). We

used a series of paired samples t-tests to compare the mean frequency power or coherence value for all subjects at each sample f_j, where j = {1,2,…,J} and J = the number of samples in our frequency series. Each test compared the mean values in our frequency series for Fa versus Ba or Ga versus Ba pairs such that, for each test, n=22 each for each trial type. This resulted in a series of p-values associated with each frequency sample 1 through J in our data set.

Because we are running a series of statistical tests, and we do not know the dependence of one sample relative to the next sample, a post-hoc correction (e.g. Bonferroni correction) was not practical. Instead, we used a Monte Carlo technique to estimate the chance-related effect size and false positive rates of this test. We scrambled the trial type assignment for each of the mean data traces in our dataset then ran the same analysis as before. We repeated this scramble procedure 10 times to determine the expected range of error that would be produced for a random set of trial type assignments. This provided an estimate of how our results should look if there was no relationship between the trial type and the acoustic recordings.

4.3 Results

4.3.1 Behavioral report

Subject performance on the three-alternative forced choice task was generally quite "accurate" in that they responded with the expected fused percept – i.e. a keypad response of "Ba" for Ba trials (98% ± 2% correct), "Da" for Ga trials (94% ± 6% correct, with one outlier who performed at only 8% accuracy [subject HME045L]), and "Va" for Fa trials (97% ± 2% correct, with the one outlier performing at 60% accuracy). During an exit interview, all subjects confirmed a strong and consistent visual capture of the sound stimulus for Fa trials, and all but the one outlier subject reported regular fusion in the Ga trials. In all cases, the perceived phonemes were consistent with previous reports, including the outlier subject. This subject reported hearing /va/ distinctly for Fa trials when the auditory and visual cues fused, something indistinct but not /ba/ for "fused" Ga trials, and a somewhat "strange and indistinct" /ba/ – close to /vba/ of /dba/ – for all trials that did not fuse.

4.3.2 Analysis of consonant articulation

We used a FFT to examine the frequency content of the periods of consonant articulation for all trial types to determine if the incongruent trials had different frequency content than the congruent trials. Fa trials have more power across the tested frequency range (0.1kHz < f < 4kHz) than Ba trials (t-test, $p<0.02$; figure 2a, blue trace). Similarly, the more restricted range around the frequency of the voiced fundamental (100Hz < f < 300Hz range; for reference, the voiced fundamental of our recorded /ba/ stimulus is approximately 130Hz) is also significantly greater for Fa as compared to Ba trials (t-test, $p<0.01$). There were, however, no significant differences between Ga and Ba trials for these same frequency ranges (t-test, $p>0.2$; figure 22a, red trace). Individual subject data were consistent with the population results; interestingly, though, subjects with greater Ba than Fa (figure 22b, red traces) frequency power also tended in that direction for Ba versus Ga trials (figure 22c, red traces) while subjects with greater Fa than Ba power (figure 22b, blue traces) tended to have a less predictable Ga versus Ba difference.

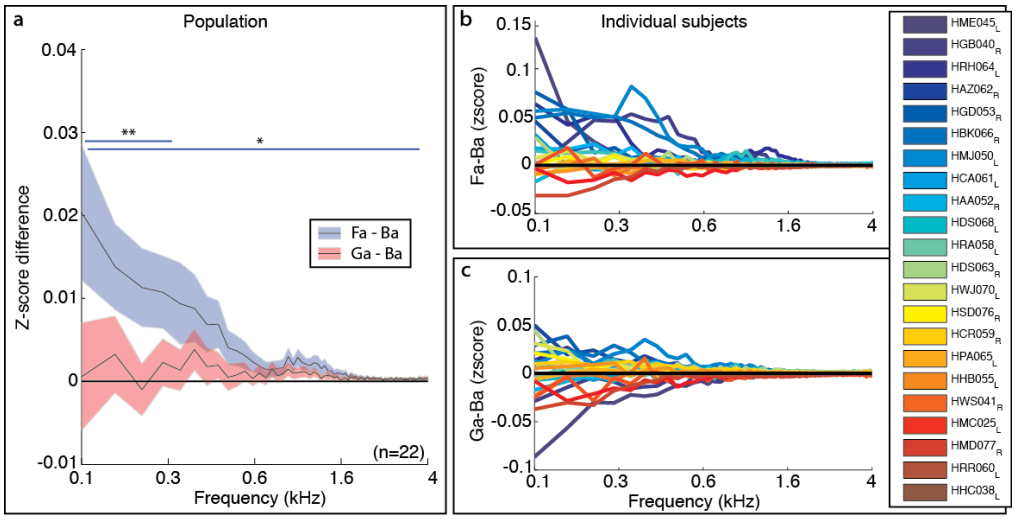

Figure 22: Difference in frequency content during consonant articulation.

(a) Population (n=22) mean ± s.e.m. difference (Fa – Ba, blue; Ga – Ba, red) in FFT values shows that the frequency power in Fa trials is significantly greater than Ba trials across the entire analyzed frequency range of 0.1 to 4kHz, particularly in the low frequencies around the fundamental pitch range of human speech (100 to 300Hz). The same is not true for Ga – Ba trials, which are not significant at either frequency range. Individual subject data for Fa – Ba (b) and Ga – Ba (c) differences are consistent with the population level differences.

The coherence between two signals is the amount of frequency power shared by both signals at a given frequency, where $0 \leq C(x,y) \leq 1$. We expect that both incongruent trial types (Fa and Ga) will have less in common with the original stimulus than the congruent (Ba) trials. We therefore calculated this measure for all trial types against the original raw stimulus and found that Fa trials had less in common with the original

stimulus than Ba trials (one-tail t-test, $p<0.01$; figure 23a, upper panel), which is particularly clear at the higher frequencies. The Ga trials follow a similar trend in that they are less similar to the original stimulus than Ba trials at higher frequencies (on-tail t-test, $p<0.05$; figure 23a, lower panel). The lack of difference between trial types at the low frequencies may be due in part to the particularly low coherence all trial types had at low frequencies (for $f < 300$ Hz, $C < 0.1$; figure 23b, lower inset), though even at higher frequencies the coherence was still fairly low ($C < 0.2$; figure 23b, upper inset). Coherence may be limited by a number of factors, including both "active" mechanisms (e.g. introduction of sound energy into the signal by active mechanisms within the ear, or changes in the acoustic impedance of the ear due to cognitive control processes) or "passive" mechanisms (e.g. static acoustic impedance levels in the ear, and the transfer function of the electronic hardware). These results indicate that, while all trial types are affected by passive alterations to the sound stimulus when it is played to when it is recorded, the "input" is more heavily altered (presumably by active mechanisms) for Fa trials than for Ba trials, and perhaps to a lesser extent with Ga as compared to Ba trials.

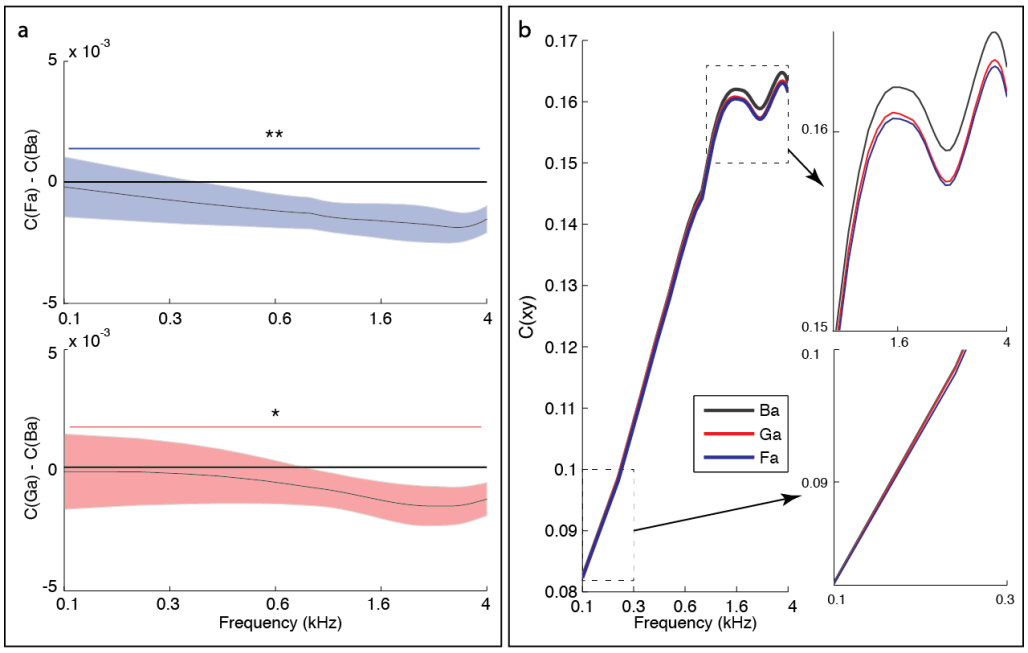

Figure 23: Differences in consonant coherence measures.

Difference (a) in coherence to original stimulus for C(Fa) – C(Ba) (a, upper panel) and C(Ga) – C(Ba) (a, lower panel). Coherence measures the shared frequency content between two signals; this shows that both Fa and Ga trials have bare less resemblance to the original stimulus than the Ba trials. (b) Mean raw coherence values at low frequency range (f < 300Hz) for all trial types is similarly poor (C(Ba) ≈ C(Fa) ≈ C(Ga); lower inset), while the differences become more apparent when coherence scores are somewhat stronger (C>0.15; upper inset).

In order to assess the degree of difference between both incongruent trial types relative to Ba trials, we ran a series of t-tests at each time point and compared the results to a Monte Carlo simulation with scrambled trial type assignments (figure 24; see

methods for details). The results of this analysis were generally similar to our initial assessment. Specifically, Fa versus Ba comparisons tended to be significantly different, particularly on the lower range of the FFT comparison (figure 24a) and upper range of the coherence comparison (figure 24b); Ga trials also had different coherence than Ba trials at higher frequencies (figure 24d), though the difference became significant higher than the Fa-Ba comparison. Ga-Ba FFT values (figure 24c) were generally slightly below the scrambled data mean ± s.e.m. level, but were typically well above p=0.05.

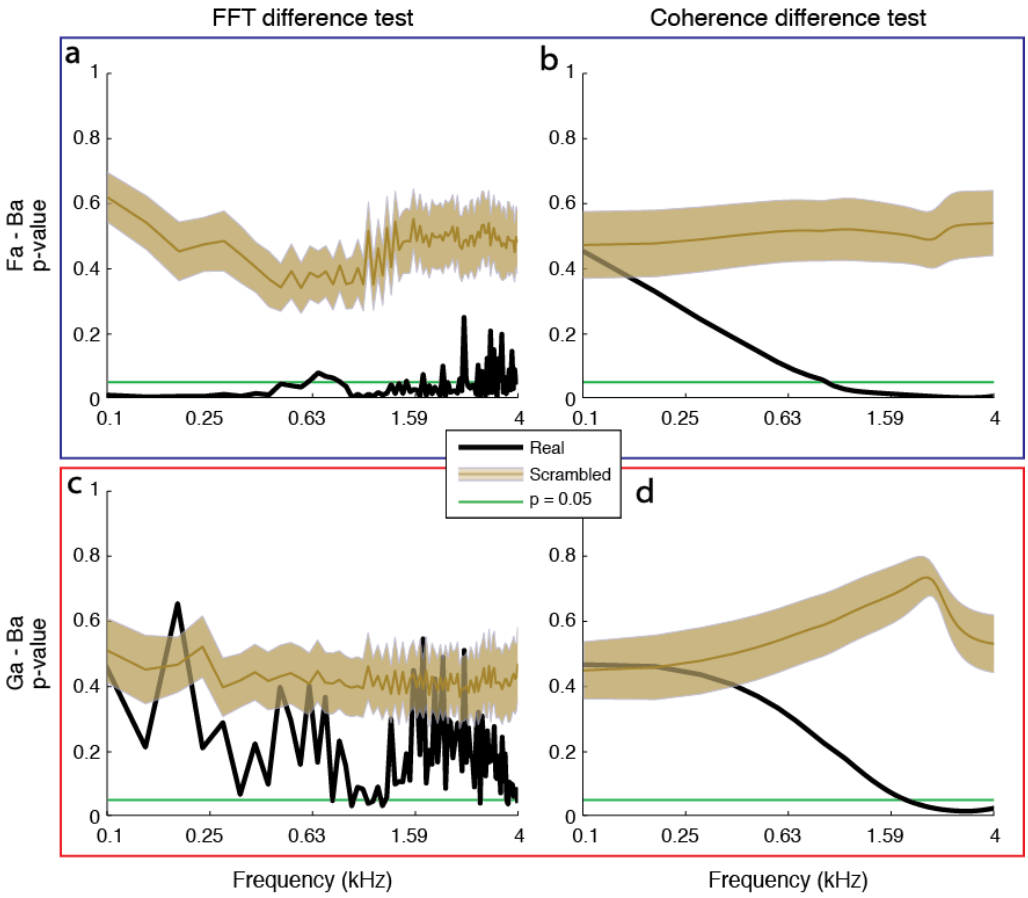

Figure 24: Comparison of consonant frequency content with t-test sequence and Monte Carlo simulation.

A more detailed profile of the frequency-based comparisons was generated by comparing the population level data for congruent Ba trials against incongruent Fa or Ga trials at each sample with a t-test (black traces; paired samples t-test, n=22). To determine an estimate of the chance error rate, the same analysis was run over 10 iterations after scrambling the trial type assignments associated with each data trace (gold trace: mean ± s.e.m.). Fa trials generally had significantly different FFT (a) and coherence (b) scores compared to Ba trials, particularly for

frequencies below 1.6 kHz for FFT analyses and above 1 kHz for the coherence measure. Ga trials were relatively similar to Ba trials across the FFT data (c), and were different at frequencies over 1.7 kHz in the coherence measure (d).

Because the phoneme /v/ has more sound power than the plosive /b/, we expect that the amplitude of the signal perceived as /v/ might be greater than perceived as /b/. In contrast, we expect that, because both /b/ and /d/ are plosive phonemes with little sound generated during the consonant, they should have similar RMS scores. Accordingly, we examined the root mean squared (RMS) amplitude difference for Fa – Ba trials and Ga – Ba trials (figure 25). We found that Fa trial do indeed have a greater amplitude than Ba trials (two-sided t-test, $p<0.01$); the same is not the case for Ga – Ba trials (t-test, $p>0.3$), which is consistent with Ga trials being perceived as a plosive /d/ consonant.

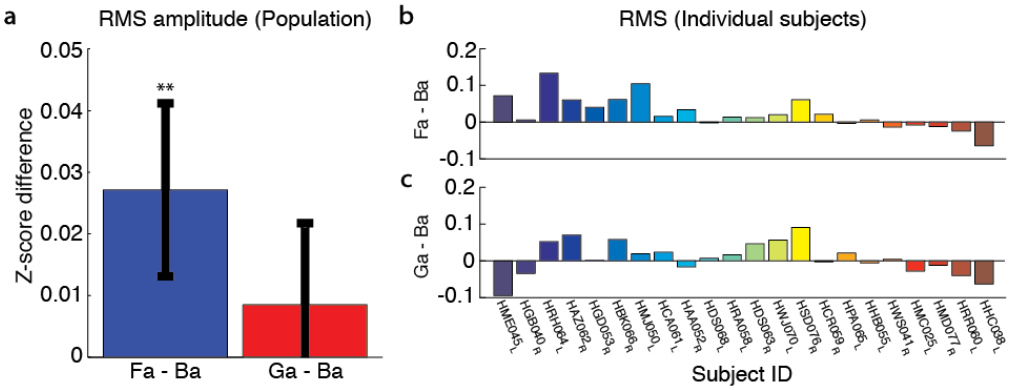

Figure 25: Difference in root mean squared (RMS) amplitude for Fa – Ba and Ga – Ba trials.

At the population level (a), Fa trials had a greater RMS amplitude than Ba type trials, in the same way that the perceived consonant /v/ has greater amplitude than the plosive phoneme /b/. This same trend is seen in 15 out of 22 individual subjects (b). Ga trials, meanwhile, were statistically similar in RMS amplitude to Ba trials. This again reflects the similarity in the perceived phonemes /b/ and /d/, which are both plosives with little sound power during articulation. Interestingly, subjects HME045L and HGB040R, who had the strongest effect at the individual level for Fa trials but for whom Ga RMS was greater than Ba, also were not particularly susceptible to McGurk fusion in the Ga trials (subject HME045L reported rarely hearing a fused percept at all, typically hearing /ba/, while subject HGB040R reported regular fusion but with an inconsistent and fluctuating percept throughout testing).

By similar reasoning to that of the RMS measure, we also expect that the variance differences between Fa – Ba trials and Ga – Ba trials will look like the RMS amplitude, and this is the case (figure 26). Fa trials had greater variance than Ba trials (t-test, $p<0.05$) while Ga trials had similar variance to Ba trials (t-test, $p>0.3$).

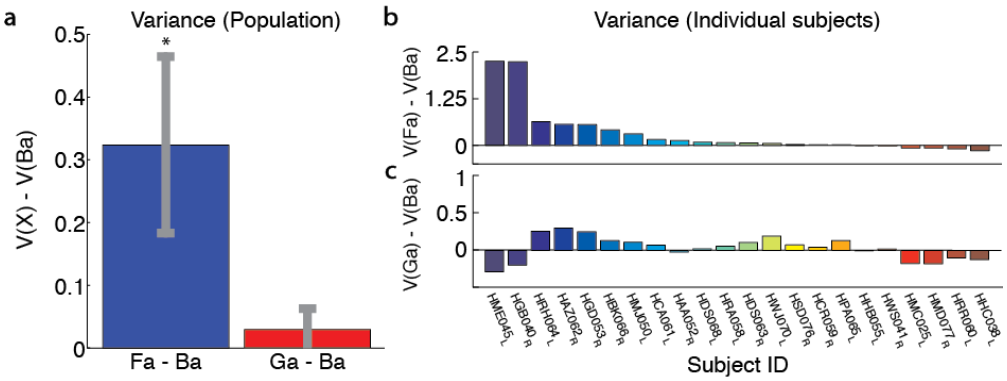

Figure 26: Difference in trial variance for Fa – Ba and Ga – Ba comparisons.

Fa trials had greater variance at the population level (a) than Ba type trials, while Ga and Ba had similar levels of variance. This same trend in Fa – Ba variance differences is seen in 16 out of 22 individual subjects (b). Similar to the RMS amplitude measure, subjects HME045L and HGB040R showed a reversal of their effect direction from Fa – Ba to Ga – Ba trials.

4.3.3 Comparison of baseline and vowel signals

We ran all of these same tests using the pre- and post-sound baseline periods to see if the differences we found between consonants were also apparent in the baseline periods. In all cases, we found no significant differences for any of the same measures as what we looked at for the consonants ($p>0.2$ in all cases; figure 27).

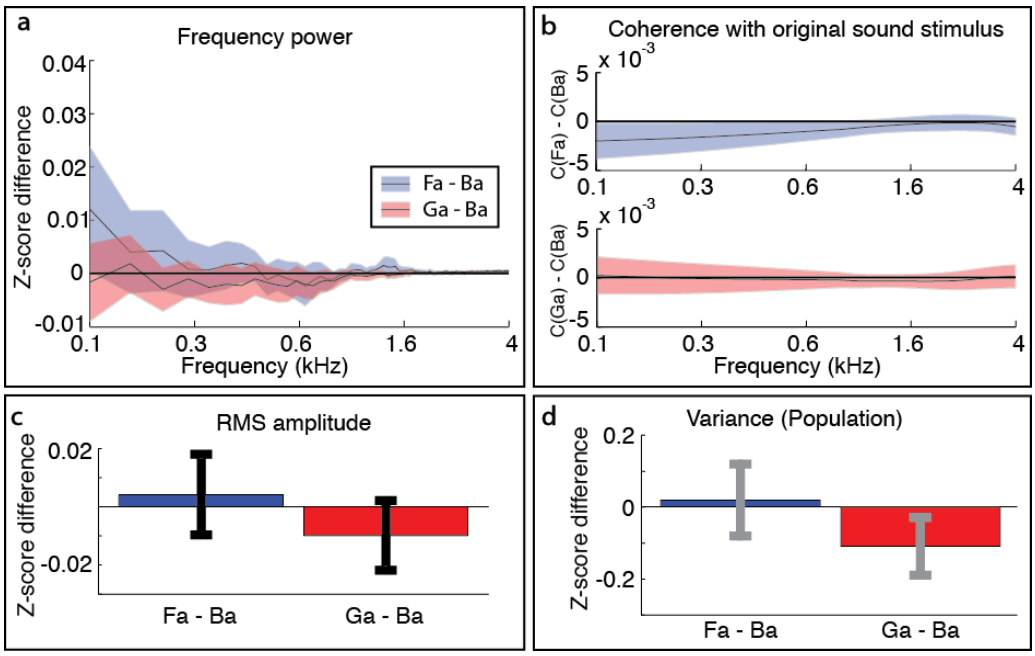

Figure 27: Baseline comparisons (all measures).

Population (n=22) statistics calculated the same way as in figure 2, except with arbitrary baseline and post-base periods instead of consonant articulations. The timing of the pre- and post-

base periods was matched in duration to the consonant duration. No significant measures were found with any of these results ($p>0.2$ in all cases).

We also examined the vowel periods in the same way and similarly found no significant differences in RMS or variance. There were differences in the frequency content for Fa – Ba and Ga – Ba trials (not significant by our measures; figure 28); however, because the vowels were all ostensibly perceived the same way, it is unclear how to interpret these differences. In both incongruent cases, there are positive differences (i.e., Fa > Ba and Ga > Ba) around the frequencies of first and second vowel formants in the original sound stimulus (f1 = 670 Hz, black arrows; f2 = 1160 Hz, gray arrows). This seems to reflect specific alterations to the frequency content of the vowel, and may be due to the interaction of the perceived consonant with the subsequent vowel sound. Intriguingly, however, this supports the overall premise of the auditory periphery filtering incoming sounds according to trial type and suggests that vowels – which are susceptible to McGurk illusion, but weakly so (Green & Gerdeman, 1995; Summerfield & McGrath, 1984) – are received differently depending on the speech context.

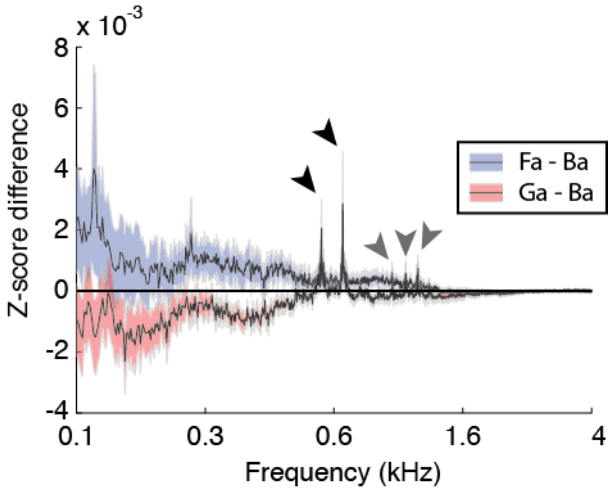

Figure 28: Differences in frequency content during vowel steady-state.

While there is no obvious hypothesis about how these sounds should differ from one another, there do appear to be differences in frequency content between congruent and incongruent trials. In particular, there appear to be differences around the frequencies of the first (black arrows) and second (gray arrows) vowel formants in the original sound stimulus, which occurred at f1 = 670 Hz and f2 = 1160 Hz.

4.4 Discussion

The auditory periphery has active mechanisms that alter sensitivity to incoming auditory stimuli. Our present results suggest that during lip-reading, the auditory periphery modifies its responsiveness in a time- and frequency-specific fashion in a manner that may contribute to perception of the specific phonemes of speech. The signals we recorded in the ear canal during presentation of the same sound varied

depending on the accompanying visual lip movement stimuli. Visual "Fa" elicited changes in the auditory signal that were consistent with the sound of "Fa", whereas visual "Ga" elicited smaller changes, consistent with the smaller differences between auditory "Ga" and auditory "Ba". Additionally, despite being difficult to interpret, the differences in the vowel steady-state data suggest that there is a subtlety to the filter properties that is not only specific to particular speech sounds in isolation, but also to the interaction of consonant-vowel pairs in a more complex manner.

This finding extends the mounting evidence that visual-related signals relevant to auditory processing can modify the actions of the auditory periphery. We have recently shown that eye movements elicit an oscillation of the eardrum that is tightly coordinated to the eye movement in time and space (Chapter II of this document). In addition, visual cues that set an expectation for incoming auditory cues produce changes in peripheral processing (Chapter III of this document; see also: Musacchia, Sams, Nicol, & Kraus, 2006; D. W. Smith et al., 2012; Srinivasan et al., 2014; Srinivasan et al., 2012). We now expand on this idea by demonstrating a highly nuanced filter system that responds to dynamic visual cues. This is particularly important for human speech, where vision is a valuable source of information that can complement ambiguous auditory information, a situation regularly encountered in our daily lives and as we age. Because proper understanding of speech requires a quick and accurate detection of subtle, rapid, and

noisy changes to the input, it is not surprising that the system would utilize any and all available information to optimize the accurate detection of these speech cues that differ from phoneme to phoneme.

We suggest that visual speech cues help the auditory periphery filter incoming auditory information by activating an initial interpretation of the possible upcoming auditory inputs. This is likely to begin in regions of the cortex thought to be involved in storing lexical and phonetic information, many of which closely overlap with audiovisual speech processing regions (Campbell, 2008). From there, such a forward model of incoming sound could be transmitted to the auditory periphery via the rich set of pathways by which signals descend from more central locations.

There are several active mechanisms in the auditory periphery that may help to filter auditory inputs based on visual cues and which may act separately or in tandem with each other. The outer hair cells (OHCs) are well known to help filter incoming sounds at the level of the basilar membrane (Fettiplace & Hackney, 2006). The middle ear muscles act on the entirety of the auditory signal as it is transmitted along the ossicular chain, and although they are not generally assumed to yield frequency specific activity, they may afford some ability to band-pass filter within a frequency range (see Chapter III of this document). While MEMs have a slower latency in response to

auditory signals than OHCs, this may not be relevant as the source of the effect must originate within the visual system and the latency with which these systems might be affected by visual input has not been measured. In fact, both systems have been shown to activate in response to self-generated sounds, including speech, and may be triggered as part of the motor-processing loop that has been hypothesized to take part in speech comprehension (also, possible sub-cortical circuit: Gruters & Groh, 2012; Hickok, 2012; Hickok, Houde, & Rong, 2011).

Perceptually, the McGurk effect is a uniquely powerful illusion: subjects who are susceptible experience the illusion even when they have been made aware that the visual information is erroneous. The power of this illusion is consistent with the idea that the bulk of the auditory system processes the incoming stimulus differently. Visually-induced changes in the periphery will ramify throughout the remainder of the auditory pathway, and suggests that the illusion is impossible to overcome because no part of the auditory pathway has access to the "raw" sound.

4.4.1 Broader implications

While this study has largely been focused on audiovisual speech, it also demonstrates a more general principle of the auditory system: that the very periphery is

influenced by multisensory stimuli. Influence of vision on peripheral auditory processes have previously been demonstrated, but only as more general attentional gating mechanism such that peripheral auditory processes are broadly suppressed during a visual task (de Boer & Thornton, 2007; Delano et al., 2007; Ferber-Viart et al., 1995; Froehlich et al., 1993; Meric & Collet, 1992, 1994; Meric et al., 1996; Puel et al., 1988; D. W. Smith et al., 2012; D. W. Smith & Keil, 2015; Srinivasan et al., 2014; Srinivasan et al., 2012). Smith et al. (D. W. Smith et al., 2012; Srinivasan et al., 2014) have also shown that the auditory periphery selectively filters incoming information according to task demands in an auditory task. Here we show that this same subtlety appears to apply in an explicit audiovisual task, where cues about the specifics of the expected auditory stimulus are embedded in the visual cue. Intriguingly, this shows that, at least in the auditory domain, the entire processing stream is necessarily multimodal and that multisensory processes are distributed throughout the entire system.

Furthermore, these data suggest that speech detection technology and auditory assistive devises such as cochlear implants and hearing aids are likely to benefit from incorporating multisensory detection and processing methods to help filter sounds at the periphery. There has been some success in developing more sensitive, accurate, and rapid artificial sensory systems through incorporating multisensory processes

(Duchnowski et al., 1995), and our data indicate that this method is biologically relevant and perhaps even optimal.

5. General conclusions

Our data show three novel findings with regard to peripheral auditory function. Specifically, the auditory periphery is sensitive to eye movements, visual cueing of auditory stimuli, and audiovisual speech. More generally, though, we have also demonstrated that the auditory system is multisensory in nature from the time that sounds first enter the ears.

Work in the auditory periphery has a reputation for being both difficult and delicate, leading to the common use of anesthesia or other techniques to mitigate the challenges associated with recording from neurons in the tiny brainstem structures of the auditory system. For example, aspiration of the visual cortex or decerebration are common techniques used to allow access to mid-brain and brain stem structures for recording but certainly disrupt the natural flow of information in the auditory system. Additionally, anesthetization of animal models has been shown to alter processing in the auditory cortex and thalamus (Szalda & Burkard, 2005; Zurita, Villa, de Ribaupierre, de Ribaupierre, & Rouiller, 1994), IC (Kuwada, Batra, & Stanford, 1989; Szalda & Burkard, 2005), and cochlear nucleus (Anderson & Young, 2004; K. Chen & Godfrey, 2000; Evans & Nelson, 1973), and the entire auditory system from periphery to cortex appears to be

impacted by behavioral state. One cannot assume that the system behaves in a normal or behaviorally relevant fashion under these sorts of conditions.

However, as more work is done with intact and awake brains, the importance of these feedback processes and cognitive influences is becoming more apparent. This series of studies serves to highlight again the impact of the efferent system on the entire auditory pathway: if these signals are changing the receptive properties of the ear itself, they are necessarily changing the entire processing stream. A complete picture of the auditory system must consider these complex interactions.

As daunting a task as this may be, the increased understanding of the efferent pathway function is particularly exciting in the field of assistive hearing devices, which behave notoriously poorly in a noisy environment. Our data are in agreement with a growing body of literature indicating that the efferent auditory pathway may be crucial to improving these devices to like-natural capabilities. Moreover, the methods used in this set of experiments offer an expansion of the already well-studied field of otoacoustic emissions. We have shown that this methodology has the capability to study more complex and nuanced effects at the auditory periphery than has previously been demonstrated.

More generally, though, these and similar results also show that multisensory interactions do not exclusively occur in later stages of auditory processing but rather are distributed throughout the entire pathway. Furthermore, these interactions are quite common, even if subtle. This in turn strongly indicates that multisensory processing is the norm – for the auditory system, at least – and that the idea of a "pure" sensory stimulus is instead the exception. Very few things occur in this world within only one sensory modality, and the brain is not equipped to process all of the information that the world provides; instead, it is exceptionally good at choosing what stimuli to ignore and, apparently, to accept.

References

Adams, J. C. (1980). Crossed and descending projections to the inferior colliculus. *Neurosci Lett, 19*(1), 1-5.

Aitkin, L., Tran, L., & Syka, J. (1994). The responses of neurons in subdivisions of the inferior colliculus of cats to tonal, noise and vocal stimuli. *Experimental Brain Research, 98*(1), 53-64.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*(3), 257-262. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14761661

Anderson, M. J., & Young, E. D. (2004). Isoflurane/N2O anesthesia suppresses narrowband but not wideband inhibition in dorsal cochlear nucleus. *Hear Res, 188*(1-2), 29-41. doi:10.1016/S0378-5955(03)00348-4

Avan, P., Loth, D., Menguy, C., & Teyssou, M. (1992). Hypothetical roles of middle ear muscles in the guinea-pig. *Hear Res, 59*(1), 59-69. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1629047

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci, 30*(7), 2414-2417. doi:10.1523/JNEUROSCI.4865-09.2010

Bergan, J. F., & Knudsen, E. I. (2009). Visual modulation of auditory responses in the owl inferior colliculus. *J Neurophysiol, 101*(6), 2924-2933. doi:10.1152/jn.91313.2008

Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: patterns of spatial generalization. *Percept Psychophys, 68*(3), 428-436. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16900834

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol Sci, 14*(6), 592-597. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14629691

Borg, E. (1972a). Excitability of the acoustic m. stapedius and m. tensor tympani reflexes in the nonanesthetized rabbit. *Acta Physiol Scand, 85*(3), 374-389. doi:10.1111/j.1748-1716.1972.tb05272.x

Borg, E. (1972b). Regulation of middle ear sound transmission in the nonanesthetized rabbit. *Acta Physiol Scand, 86*(2), 175-190. doi:10.1111/j.1748-1716.1972.tb05324.x

Borg, E., & Moller, A. R. (1968). The acoustic middle ear reflex in unanesthetized rabbits. *Acta Otolaryngol, 65*(6), 575-585. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5706028

Brainard, M. S., & Knudsen, E. I. (1993). Visual calibration of the neural representation of auditory space in the barn owl. *Biomedical Research, 14*(SUPPL. 4), 35-40.

Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci, 25*(29), 6797-6806. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16033889

Buchan, J. N., Pare, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Soc Neurosci, 2*(1), 1-13. doi:10.1080/17470910601043644

Bulkin, D. A., & Groh, J. M. (2012a). Distribution of eye position information in the monkey inferior colliculus. *J Neurophysiol, 107*(3), 785-795. doi:10.1152/jn.00662.2011

Bulkin, D. A., & Groh, J. M. (2012b). Distribution of visual and saccade related information in the monkey inferior colliculus. *Front Neural Circuits, 6*, 61. doi:10.3389/fncir.2012.00061

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proc Natl Acad Sci U S A, 102*(51), 18751-18756. doi:10.1073/pnas.0507704102

Busse, L., & Woldorff, M. G. (2003). The ERP omitted stimulus response to "no-stim" events and its implications for fast-rate event-related fMRI designs. *Neuroimage, 18*(4), 856-864. doi:10.1016/s1053-8119(03)00012-0

Cai, T. T., Liang, T., & Zhou, H. H. (2013). Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. *arXiv preprint*. doi:arXiv:1309.0482.

Calford, M. B., & Aitkin, L. M. (1983). Ascending projections to the medial geniculate body of the cat: evidence for multiple, parallel auditory pathways through thalamus. *Journal Of Neuroscience, 3*(11), 2365-2380.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philos Trans R Soc Lond B Biol Sci, 363*(1493), 1001-1010. doi:10.1098/rstb.2007.2155

Carmel, P. W., & Starr, A. (1963). Acoustic and nonacoustic factors modifying middle-ear muscle activity in waking cats. *J Neurophysiol, 26*, 598-616. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14018722

Carpenter, M. B. (1959). Lesions of the fastigial nuclei in the rhesus monkey. *Am J Anat, 104*, 1-33. doi:10.1002/aja.1001040102

Casselbrant, M., Ingelstedt, S., & Ivarsson, A. (1977). Volume displacement of the tympanic membrane in the sitting position as a function of middle ear muscle activity. A quantitative microflow method. *Acta Otolaryngol, 84*(5-6), 402-413. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/144403

Champoux, F., Paiement, P., Mercier, C., Lepore, F., Lassonde, M., & Gagne, J. P. (2007). Auditory processing in a patient with a unilateral lesion of the inferior colliculus. *Eur J Neurosci, 25*(1), 291-297. doi:10.1111/j.1460-9568.2006.05260.x

Champoux, F., Tremblay, C., Mercier, C., Lassonde, M., Lepore, F., Gagne, J. P., & Theoret, H. (2006). A role for the inferior colliculus in multisensory speech integration. *NeuroReport, 17*(15), 1607-1610. doi:10.1097/01.wnr.0000236856.93586.94

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput Biol, 5*(7), e1000436. doi:10.1371/journal.pcbi.1000436

Chen, K., & Godfrey, D. A. (2000). Sodium pentobarbital abolishes bursting spontaneous activity of dorsal cochlear nucleus in rat brain slices. *Hear Res, 149*(1-2), 216-222. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11033260

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Atten Percept Psychophys, 75*(5), 790-811. doi:10.3758/s13414-013-0475-4

Cheng, J. T., Aarnisalo, A. A., Harrington, E., Hernandez-Montes Mdel, S., Furlong, C., Merchant, S. N., & Rosowski, J. J. (2010). Motion of the surface of the human tympanic membrane measured with stroboscopic holography. *Hear Res, 263*(1-2), 66-77. doi:10.1016/j.heares.2009.12.024

Chow, K. L., & Hutt, P. J. (1953). The association cortex of Macaca mulatta: a review of recent contributions to its anatomy and functions. *Brain, 76*(4), 625-677. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13115577

Coleman, J. R., & Clerici, W. J. (1987). Sources of projections to subdivisions of the inferior colliculus in the rat. *J Comp Neurol, 262*(2), 215-226. doi:10.1002/cne.902620204

Cooper, M. H., & Young, P. A. (1976). Cortical projections to the inferior colliculus of the cat. *Exp Neurol, 51*(2), 488-502. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1269575

Covey, E., Hall, W. C., & Kobler, J. B. (1987). Subcortical connections of the superior colliculus in the mustache bat, Pteronotus parnellii. *J Comp Neurol, 263*(2), 179-197. doi:10.1002/cne.902630203

de Boer, J., & Thornton, A. R. (2007). Effect of subject task on contralateral suppression of click evoked otoacoustic emissions. *Hear Res, 233*(1-2), 117-123. doi:10.1016/j.heares.2007.08.002

De Gennaro, L., & Ferrara, M. (2000). Sleep deprivation and phasic activity of REM sleep: independence of middle-ear muscle activity from rapid eye movements. *Sleep, 23*(1), 81-85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10678468

De Gennaro, L., Ferrara, M., Urbani, L., & Bertini, M. (2000). A complementary relationship between wake and REM sleep in the auditory system: a pre-sleep increase of middle-ear muscle activity (MEMA) causes a decrease of MEMA during sleep. *Exp Brain Res, 130*(1), 105-112. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10638447

De Greef, D., Aernouts, J., Aerts, J., Cheng, J. T., Horwitz, R., Rosowski, J. J., & Dirckx, J. J. (2014). Viscoelastic properties of the human tympanic membrane studied with stroboscopic holography and finite element modeling. *Hear Res, 312*, 69-80. doi:10.1016/j.heares.2014.03.002

Delano, P. H., Elgueda, D., Hamame, C. M., & Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *J Neurosci, 27*(15), 4146-4153. doi:10.1523/JNEUROSCI.3702-06.2007

Dewson, J. H., 3rd, Dement, W. C., & Simmons, F. B. (1965). Middle Ear Muscle Activity in Cats during Sleep. *Exp Neurol, 12*, 1-8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14285351

Djupesland, G. (1964). Middle Ear Muscle Reflexes Elicited by Acoustic and Nonacoustic Stimulation. *Acta Otolaryngol Suppl, 188*, SUPPL 188:287+. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14146687

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Front Integr Neurosci, 9*, 32. doi:10.3389/fnint.2015.00032

Donohue, S. E., Roberts, K. C., Grent-'t-Jong, T., & Woldorff, M. G. (2011). The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. *J Neurosci, 31*(22), 7982-7990. doi:10.1523/JNEUROSCI.5298-10.2011

Drager, U. C., & Hubel, D. H. (1975). Physiology of visual cells in mouse superior colliculus and correlation with somatosensory and auditory input. *Nature, 253*, 203-204.

Druga, R., & Syka, J. (1984). Projections from auditory structures to the superior colliculus in the rat. *Neurosci Lett, 45*(3), 247-252. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6728318

Duchnowski, P., Hunke, M., Busching, D., Meier, U., & Waibel, A. (1995). Toward Movement-Invariant Automatic Lip-Reading and Speech Recognition. *1995 International Conference on Acoustics, Speech, and Signal Processing - Conference Proceedings, Vols 1-5*, 109-112. Retrieved from <Go to ISI>://WOS:A1995BD41X00028

Earle, A. M., & Matzke, H. A. (1974). Efferent fibers of the deep cerebellar nuclei in hedgehogs. *J Comp Neurol, 154*(2), 117-131. doi:10.1002/cne.901540202

Edwards, T. J. (1981). Multiple Features Analysis of Intervocalic English Plosives. *Journal of the Acoustical Society of America, 69*(2), 535-547. doi:Doi 10.1121/1.385482

Eliasson, S., & Gisselsson, L. (1955). Electromyographic studies of the middle ear muscles of the cat. *Electroencephalogr Clin Neurophysiol, 7*(3), 399-406. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13251182

Evans, E. F., & Nelson, P. G. (1973). The responses of single neurones in the cochlear nucleus of the cat as a function of their location and the anaesthetic state. *Exp Brain Res, 17*(4), 402-427. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4725899

Ferber-Viart, C., Duclaux, R., Collet, L., & Guyonnard, F. (1995). Influence of auditory stimulation and visual attention on otoacoustic emissions. *Physiol Behav, 57*(6), 1075-1079. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7652027

Fettiplace, R., & Hackney, C. M. (2006). The sensory and motor roles of auditory hair cells. *Nat Rev Neurosci, 7*(1), 19-29. doi:10.1038/nrn1828

Fischer, C., Bognar, L., Turjman, F., & Lapras, C. (1995). Auditory evoked potentials in a patient with a unilateral lesion of the inferior colliculus and medial geniculate body. *Electroencephalography & Clinical Neurophysiology, 96*(3), 261-267.

Freedman, E. G., Stanford, T. R., & Sparks, D. L. (1996). Combined eye-head gaze shifts produced by electrical stimulation of the superior colliculus in rhesus monkeys. *J Neurophysiol, 76*(2), 927-952. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=8871209

Frissen, I., Vroomen, J., & de Gelder, B. (2012). The aftereffects of ventriloquism: the time course of the visual recalibration of auditory localization. *Seeing Perceiving, 25*(1), 1-14. doi:10.1163/187847611X620883

Froehlich, P., Collet, L., & Morgon, A. (1993). Transiently evoked otoacoustic emission amplitudes change with changes of directed attention. *Physiol Behav, 53*(4), 679-682. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8511172

Fu, K. M., Shah, A. S., O'Connell, M. N., McGinnis, T., Eckholdt, H., Lakatos, P., . . . Schroeder, C. E. (2004). Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. *J Neurophysiol, 92*(6), 3522-3531. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15282263

Gandhi, N. J., & Katnani, H. A. (2011). Motor functions of the superior colliculus. *Annu Rev Neurosci, 34*, 205-231. doi:10.1146/annurev-neuro-061010-113728

Gelfand, S. A. (1984). The contralateral acoustic reflex. In S. Silman (Ed.), *The acoustic reflex: Basic principles and clinical applications* (pp. 137-186). New York, NY: Marcel Dekker.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci, 25*(20), 5004-5012. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15901781

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn Sci, 10*(6), 278-285. doi:10.1016/j.tics.2006.04.008

Gordon, B. (1973). Receptive fields in deep layers of cat superior colliculus. *J Neurophysiol., 36*(2), 157-178.

Green, K. P., & Gerdeman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech information: the McGurk effect with mismatched vowels. *J Exp Psychol Hum Percept Perform, 21*(6), 1409-1426. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7490588

Greisen, O., & Neergaard, E. B. (1975). Middle ear reflex activity in the startle reaction. *Arch Otolaryngol, 101*(6), 348-353. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1131100

Groh, J. M., & Pai, D. (2010). Looking at sounds: neural mechanisms in the primate brain. In M. L. Platt & A. A. Ghazanfar (Eds.), *Primate Neuroethology*. New York, NY: Oxford University Press.

Groh, J. M., & Sparks, D. L. (1992). Two models for transforming auditory signals from head-centered to eye- centered coordinates. *Biol Cybern, 67*(4), 291-302. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=1515508

Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye position influences auditory responses in primate inferior colliculus. *Neuron, 29*(2), 509-518. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=11239439

Gruters, K. G., & Groh, J. M. (2012). Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Front Neural Circuits, 6*, 96. doi:10.3389/fncir.2012.00096

Guinan, J. J., Jr. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear Hear, 27*(6), 589-607. doi:10.1097/01.aud.0000240507.83072.e7

Guinan, J. J., Jr. (2010). Cochlear efferent innervation and function. *Curr Opin Otolaryngol Head Neck Surg, 18*(5), 447-453. doi:10.1097/MOO.0b013e32833e05d6

Gutfreund, Y., Zheng, W., & Knudsen, E. I. (2002). Gated visual input to the central auditory system. *Science, 297*(5586), 1556-1559. doi:10.1126/science.1073712

Harting, J. K., & Van Lieshout, D. P. (2000). Projections from the rostral pole of the inferior colliculus to the cat superior colliculus. *Brain Res, 881*(2), 244-247. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11036169

Herbert, H., Klepper, A., & Ostwald, J. (1997). Afferent and efferent connections of the ventrolateral tegmental area in the rat. *Anat Embryol (Berl), 196*(3), 235-259. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9310315

Herbin, M., Reperant, J., & Cooper, H. M. (1994). Visual system of the fossorial mole-lemmings, Ellobius talpinus and Ellobius lutescens. *J Comp Neurol, 346*(2), 253-275. doi:10.1002/cne.903460206

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat Rev Neurosci, 13*(2), 135-145. doi:10.1038/nrn3158

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron, 69*(3), 407-422. doi:10.1016/j.neuron.2011.01.019

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am, 97*(5 Pt 1), 3099-3111. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7759650

Holst, H. E., Ingelstedt, S., & Ortegren, U. (1963). Ear drum movements following stimulation of the middle ear muscles. *Acta Otolaryngol Suppl, 182*, 73-89. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13961437

Honrubia, F. M., & Elliott, J. H. (1968). Efferent innervation of the retina. I. Morphologic study of the human retina. *Arch Ophthalmol, 80*(1), 98-103. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4873166

Honrubia, F. M., & Elliott, J. H. (1970). Efferent innervation of the retina. II. Morphologic study of the monkey retina. *Invest Ophthalmol, 9*(12), 971-976. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4098604

Hopkins, D. A., & Holstege, G. (1978). Amygdaloid projections to the mesencephalon, pons and medulla oblongata in the cat. *Exp Brain Res, 32*(4), 529-547. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/689127

Huffman, R. F., & Henson, O. W. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Research Reviews, 15*, 295-323.

Hugelin, A., Dumont, S., & Paillas, N. (1960). Tympanic muscles and control of auditory input during arousal. *Science, 131*(3410), 1371-1372. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14403819

Huttenbrink, K. B. (1988). The mechanics of the middle-ear at static air pressures: the role of the ossicular joints, the function of the middle-ear muscles and the behaviour of stapedial prostheses. *Acta Otolaryngol Suppl, 451*, 1-35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3218485

Huttenbrink, K. B. (1989). [Movement of the ear ossicles by middle ear muscle contraction]. *Laryngorhinootologie, 68*(11), 614-621. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2604816

Hyde, P. S., & Knudsen, E. I. (2000). Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *J Comp Neurol, 421*(2), 146-160. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10813778

Itaya, S. K., & Itaya, P. W. (1985). Centrifugal fibers to the rat retina from the medial pretectal area and the periaqueductal grey matter. *Brain Res, 326*(2), 362-365. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=3971160

Itaya, S. K., & Van Hoesen, G. W. (1982). Retinal innervation of the inferior colliculus in rat and monkey. *Brain Res, 233*(1), 45-52.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept Mot Skills, 37*(3), 967-979. doi:10.2466/pms.1973.37.3.967

Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature, 309*(5966), 345-347. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=6727988

Judge, S. J., Richmond, B. J., & Chu, F. C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Res., 20*, 535-538.

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *J Acoust Soc Am, 64*(5), 1386-1391. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/744838

Kemp, D. T. (1979). Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea. *Arch Otorhinolaryngol, 224*(1-2), 37-45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/485948

Kemp, D. T., Ryan, S., & Bray, P. (1990). A guide to the effective use of otoacoustic emissions. *Ear Hear, 11*(2), 93-105. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2340969

Khaleghi, M., Furlong, C., Ravicz, M., Cheng, J. T., & Rosowski, J. J. (2015). Three-dimensional vibrometry of the human eardrum with stroboscopic lensless digital holography. *J Biomed Opt, 20*(5), 051028. doi:10.1117/1.JBO.20.5.051028

Klug, A., Bauer, E. E., Hanson, J. T., Hurley, L., Meitzen, J., & Pollak, G. D. (2002). Response selectivity for species-specific calls in the inferior colliculus of Mexican free-tailed bats is generated by inhibition. *J Neurophysiol, 88*(4), 1941-1954. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12364520

Kopco, N., Lin, I. F., Shinn-Cunningham, B. G., & Groh, J. M. (2009). Reference frame of the ventriloquism aftereffect. *J Neurosci, 29*(44), 13809-13814. doi:10.1523/JNEUROSCI.2783-09.2009

Kudo, M., & Niimi, K. (1980). Ascending projections of the inferior colliculus in the cat: an autoradiographic study. *J Comp Neurol, 191*(4), 545-556. doi:10.1002/cne.901910403

Kunimoto, Y., Hasegawa, K., Arii, S., Kataoka, H., Yazama, H., Kuya, J., & Kitano, H. (2014). Sequential multipoint motion of the tympanic membrane measured by laser Doppler vibrometry: preliminary results for normal tympanic membrane. *Otol Neurotol, 35*(4), 719-724. doi:10.1097/MAO.0000000000000242

Kuwada, S., Batra, R., & Stanford, T. R. (1989). Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit: effects of sodium pentobarbital. *J Neurophysiol, 61*(2), 269-282. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=2918355

Lee, J., & Groh, J. M. (2009). *Eye-centered reference frame of auditory and visual oculomotor signals in the primate superior colliculus.* Paper presented at the Soc. Neurosci. Abstr.

Lee, J., & Groh, J. M. (2012). Auditory signals evolve from hybrid- to eye-centered coordinates in the primate superior colliculus. *J Neurophysiol*. doi:10.1152/jn.00706.2011

MacDonald, J., & McGurk, H. (1978a). Visual influences on speech perception processes. *Perception & Psychophysics, 24*(3), 253-257.

MacDonald, J., & McGurk, H. (1978b). Visual influences on speech perception processes. *Percept Psychophys, 24*(3), 253-257. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/704285

Maier, J. X., & Groh, J. M. (2010). Comparison of gain-like properties of eye position signals in inferior colliculus versus auditory cortex of primates. *Front Integr Neurosci, 4*. doi:10.3389/fnint.2010.00121

Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *J Acoust Soc Am, 125*(6), 3962-3973. doi:10.1121/1.2990715

Marsh, R. A., Fuzessery, Z. M., Grose, C. D., & Wenstrup, J. J. (2002). Projection to the inferior colliculus from the basal nucleus of the amygdala. *J Neurosci, 22*(23), 10449-10460. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12451144

Mascetti, G. G., & Strozzi, L. (1988). Visual cells in the inferior colliculus of the cat. *Brain Res, 442*(2), 387-390.

Mast, T. E., & Chung, D. Y. (1973). Binaural interaction in the superior colliculus of the chinchilla. *Brain Res, 62*(1), 227-230. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4765113

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Meric, C., & Collet, L. (1992). Visual attention and evoked otoacoustic emissions: a slight but real effect. *Int J Psychophysiol, 12*(3), 233-235. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1639669

Meric, C., & Collet, L. (1994). Differential effects of visual attention on spontaneous and evoked otoacoustic emissions. *Int J Psychophysiol, 17*(3), 281-289. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7806471

Meric, C., Micheyl, C., & Collet, L. (1996). Attention and evoked otoacoustic emissions: attempts at characterization of intersubject variation. *Physiol Behav, 59*(1), 1-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8848467

Metzger, R. R., Greene, N. T., Porter, K. K., & Groh, J. M. (2006). Effects of reward and behavioral context on neural activity in the primate inferior colliculus. *J Neurosci, 26*(28), 7468-7476. doi:10.1523/JNEUROSCI.5401-05.2006

Metzger, R. R., Kelly, K. A., & Groh, J. M. (2004). Sensitivity to eye position in the inferior colliculus of the monkey during an auditory saccade task. *Soc Neurosci. Abstr.*

Metzger, R. R., Mullette-Gillman, O. A., Underhill, A. M., Cohen, Y. E., & Groh, J. M. (2004). Auditory saccades from different eye positions in the monkey: implications for coordinate transformations. *J Neurophysiol, 92*(4), 2622-2627. doi:10.1152/jn.00326.2004

Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional Deafness: Visual Load Leads to Time-Specific Suppression of Auditory Evoked Responses. *J Neurosci, 35*(49), 16046-16054. doi:10.1523/JNEUROSCI.2931-15.2015

Moriizumi, T., Leduc-Cross, B., Wu, J. Y., & Hattori, T. (1992). Separate neuronal populations of the rat substantia nigra pars lateralis with distinct projection sites and transmitter phenotypes. *Neuroscience, 46*(3), 711-720. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1372117

Mukerji, S., Windsor, A. M., & Lee, D. J. (2010). Auditory brainstem circuits that mediate the middle ear muscle reflex. *Trends Amplif, 14*(3), 170-191. doi:10.1177/1084713810381771

Mullette-Gillman, O. A., Cohen, Y. E., & Groh, J. M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *J Neurophysiol, 94*(4), 2331-2352. doi:10.1152/jn.00021.2005

Mullette-Gillman, O. A., Cohen, Y. E., & Groh, J. M. (2009). Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame. *Cereb Cortex, 19*(8), 1761-1775. doi:10.1093/cercor/bhn207

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Exp Brain Res, 168*(1-2), 1-10. doi:10.1007/s00221-005-0071-5

Nienhuis, R., & Olds, J. (1978). Changes in unit responses to tones after food reinforcement in the auditory pathway of the rat: intertrial arousal. *Exp Neurol, 59*(2), 229-242. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/639917

Ono, T., Nishijo, H., & Nishino, H. (2000). Functional role of the limbic system and basal ganglia in motivated behaviors. *J Neurol, 247 Suppl 5*, V23-32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11081801

Pare, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: the influence of ocular fixations on the McGurk effect. *Percept Psychophys, 65*(4), 553-567. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12812278

Paula-Barbosa, M. M., & Sousa-Pinto, A. (1973). Auditory cortical projections to the superior colliculus in the cat. *Brain Res, 50*(1), 47-61.

Pessah, M. A., & Roffwarg, H. P. (1972). Spontaneous middle ear muscle activity in man: a rapid eye movement sleep phenomenon. *Science, 178*(4062), 773-776. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4343261

Pincherli Castellanos, T. A., Aitoubah, J., Molotchnikoff, S., Lepore, F., & Guillemot, J. P. (2007). Responses of inferior collicular cells to species-specific vocalizations in normal and enucleated rats. *Exp Brain Res, 183*(3), 341-350. doi:10.1007/s00221-007-1049-2

Populin, L. C., Tollin, D. J., & Yin, T. C. (2004). Effect of eye position on saccades and neuronal responses to acoustic stimuli in the superior colliculus of the behaving cat. *J Neurophysiol, 92*(4), 2151-2167. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15190094

Porter, J. D. (1986). Brainstem terminations of extraocular muscle primary afferent neurons in the monkey. *J Comp Neurol, 247*(2), 133-143. doi:10.1002/cne.902470202

Porter, K. K., & Groh, J. M. (2006). The other transformation required for visual-auditory integration: representational format. *Prog Brain Res, 155*, 313-323. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17027396

Porter, K. K., Metzger, R. R., & Groh, J. M. (2006). Representation of eye position in primate inferior colliculus. *J Neurophysiol, 95*(3), 1826-1842. doi:10.1152/jn.00857.2005

Porter, K. K., Metzger, R. R., & Groh, J. M. (2007). Visual- and saccade-related signals in the primate inferior colliculus. *Proc Natl Acad Sci U S A, 104*(45), 17855-17860. doi:10.1073/pnas.0706249104

Probst, R. (1990). Otoacoustic emissions: an overview. *Adv Otorhinolaryngol, 44*, 1-91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2407069

Puel, J. L., Bonfils, P., & Pujol, R. (1988). Selective attention modifies the active micromechanical properties of the cochlea. *Brain Res, 447*(2), 380-383. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3390709

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc Natl Acad Sci U S A, 95*(3), 869-875. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9448253

Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory selective attention modulates activation of human inferior colliculus. *J Neurophysiol, 100*(6), 3323-3327. doi:10.1152/jn.90607.2008

Robinson, D. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng., 10*, 137-145.

Romanski, L. M., & Diehl, M. M. (2011). Neurons responsive to face-view in the primate ventrolateral prefrontal cortex. *Neuroscience, 189*, 223-235. doi:10.1016/j.neuroscience.2011.05.014

Rosowski, J. J., Cheng, J. T., Merchant, S. N., Harrington, E., & Furlong, C. (2011). New data on the motion of the normal and reconstructed tympanic membrane. *Otol Neurotol, 32*(9), 1559-1567. doi:10.1097/MAO.0b013e31822e94f3

Ruth, R. E., Rosenfeld, J. P., Harris, D. M., & Birkel, P. (1974). Effects of aversive and rewarding electrical brain stimulation on auditory evoked responses in albino rat tectum. *Physiol Behav, 13*(6), 729-735. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4445279

Ryan, A., & Miller, J. (1977). Effects of behavioral performance on single-unit firing patterns in inferior colliculus of the rhesus monkey. *J Neurophysiol, 40*(4), 943-956.

Ryan, A. F., Miller, J. M., Pfingst, B. E., & Martin, G. K. (1984). Effects of reaction time performance on single-unit activity in the central auditory pathway of the rhesus macaque. *Journal Of Neuroscience, 4*(1), 298-308.

Salomon, G., & Starr, A. (1963). Electromyography of middle ear muscles in man during motor activities. *Acta Neurol Scand, 39*, 161-168. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13991171

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res, 14*(1), 187-198. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12063142

Shammah-Lagnado, S. J., Alheid, G. F., & Heimer, L. (1996). Efferent connections of the caudal part of the globus pallidus in the rat. *J Comp Neurol, 376*(3), 489-507. doi:10.1002/(SICI)1096-9861(19961216)376:3<489::AID-CNE10>3.0.CO;2-H

Shinonaga, Y., Takada, M., Ogawa-Meguro, R., Ikai, Y., & Mizuno, N. (1992). Direct projections from the globus pallidus to the midbrain and pons in the cat. *Neurosci Lett, 135*(2), 179-183. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1625791

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex, 17*(10), 2387-2399. doi:10.1093/cercor/bhl147

Smith, D. W., Aouad, R. K., & Keil, A. (2012). Cognitive task demands modulate the sensitivity of the human cochlea. *Front Psychol, 3*, 30. doi:10.3389/fpsyg.2012.00030

Smith, D. W., & Keil, A. (2015). The biological role of the medial olivocochlear efferents in hearing: separating evolved function from exaptation. *Front Syst Neurosci, 9*, 12. doi:10.3389/fnsys.2015.00012

Smith, P. F. (2012). Interactions between the vestibular nucleus and the dorsal cochlear nucleus: implications for tinnitus. *Hear Res, 292*(1-2), 80-82. doi:10.1016/j.heares.2012.08.006

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk-MacDonald illusion. *J Exp Psychol Hum Percept Perform, 35*(2), 580-587. doi:10.1037/a0013483

Srinivasan, S., Keil, A., Stratis, K., Osborne, A. F., Cerwonka, C., Wong, J., . . . Smith, D. W. (2014). Interaural attention modulates outer hair cell function. *Eur J Neurosci, 40*(12), 3785-3792. doi:10.1111/ejn.12746

Srinivasan, S., Keil, A., Stratis, K., Woodruff Carr, K. L., & Smith, D. W. (2012). Effects of cross-modal selective attention on the sensory periphery: cochlear sensitivity is altered by selective attention. *Neuroscience, 223*, 325-332. doi:10.1016/j.neuroscience.2012.07.062

Stahl, J. S. (2001). Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity. *Exp Brain Res, 136*(2), 200-210. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11206282

Steinbach, M. J. (1987). Proprioceptive knowledge of eye position. *Vision Res, 27*(10), 1737-1744.

Stitt, I., Galindo-Leon, E., Pieper, F., Hollensteiner, K. J., Engler, G., & Engel, A. K. (2015). Auditory and visual interactions between the superior and inferior colliculi in the ferret. *Eur J Neurosci, 41*(10), 1311-1320. doi:10.1111/ejn.12847

Stoica, P., & Moses, R. L. (2005). *Spectral analysis of signals*. Upper Saddle River, N.J.: Pearson/Prentice Hall.

Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of auditory and visual communication information in the primate ventrolateral

prefrontal cortex. *J Neurosci, 26*(43), 11138-11147. doi:10.1523/JNEUROSCI.3550-06.2006

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America, 26*(2), 212-215. doi:Doi 10.1121/1.1907309

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *Q J Exp Psychol A, 36*(1), 51-74. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6536037

Suta, D., Kvasnak, E., Popelar, J., & Syka, J. (2003). Representation of species-specific vocalizations in the inferior colliculus of the guinea pig. *J Neurophysiol, 90*(6), 3794-3808. doi:10.1152/jn.01175.2002

Syka, J., & Radil-Weiss, T. (1973). Acoustical responses of inferior colliculus neurons in rats influenced by sciatic nerve stimulation and light flashes. *Int J Neurosci, 5*(5), 201-206. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4781426

Szalda, K., & Burkard, R. (2005). The effects of nembutal anesthesia on the auditory steady-state response (ASSR) from the inferior colliculus and auditory cortex of the chinchilla. *Hear Res, 203*(1-2), 32-44. doi:10.1016/j.heares.2004.11.014

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci, 14*(9), 400-410. doi:10.1016/j.tics.2010.06.008

Tammer, R., Ehrenreich, L., & Jurgens, U. (2004). Telemetrically recorded neuronal activity in the inferior colliculus and bordering tegmentum during vocal communication in squirrel monkeys (Saimiri sciureus). *Behav Brain Res, 151*(1-2), 331-336. doi:10.1016/j.bbr.2003.09.008

Tawil, R. N., Saade, N. E., Bitar, M., & Jabbur, S. J. (1983). Polysensory interactions on single neurons of cat inferior colliculus. *Brain Res, 269*(1), 149-152. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6307479

Teig, E. (1972). Tension and contraction time of motor units of the middle ear muscles in the cat. *Acta Physiol Scand, 84*(1), 11-21. doi:10.1111/j.1748-1716.1972.tb05150.x

Thier, P., & Mock, M. (2005). The oculomotor role of the pontine nuclei and the nucleus reticularis tegmenti pontis. *Prog Brain Res, 151*, 293-320. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16221593

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the "ventriloquism effect". *Percept Mot Skills, 36*(3), 1171-1184. doi:10.2466/pms.1973.36.3c.1171

Thurlow, W. R., & Rosenthal, T. M. (1976). Further study of existence regions for the "ventriloquism effect". *Journal of the American Audiology Society, 1*(6), 280-286.

Van Buskirk, R. L. (1983). Subcortical auditory and somatosensory afferents to hamster superior colliculus. *Brain Res Bull, 10*(5), 583-587. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6871733

van den Berge, H., van Geest, A., Rensema, J. W., & Drukker, J. (1990). Three-dimensional graphic reconstruction of the tympanic bulla of the rat with special reference to the middle ear muscles. *Acta Otolaryngol, 110*(3-4), 253-261. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2239215

van Son, R. J. J. H., & Pols, L. C. W. (1999). An acoustic description of consonant reduction. *Speech Communication, 28*(2), 125-140. doi:Doi 10.1016/S0167-6393(99)00009-6

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychol (Amst), 108*(1), 21-33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11485191

Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept Psychophys, 63*(4), 651-659. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11436735

Vroomen, J., de Gelder, B., & Vroomen, J. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *J Exp Psychol Hum Percept Perform, 30*(3), 513-518. doi:10.1037/0096-1523.30.3.513

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J Exp Psychol Hum Percept Perform, 32*(4), 1063-1071. doi:10.1037/0096-1523.32.4.1063

Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychol (Amst), 130*(1), 58-63. doi:10.1016/j.actpsy.2008.10.001

Weiss, H. S., Mundie, J. R., Jr., Cashin, J. L., & Shinabarger, E. W. (1962). The normal human intra-aural muscle reflex in response to sound. *Acta Otolaryngol, 55*, 505-515. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13999721

Weissman, D. H., Warner, L. M., & Woldorff, M. G. (2004). The neural mechanisms for minimizing cross-modal distraction. *J Neurosci, 24*(48), 10941-10949. doi:10.1523/JNEUROSCI.3669-04.2004

Werner-Reiss, U., Kelly, K. A., Trause, A. S., Underhill, A. M., & Groh, J. M. (2003). Eye position affects activity in primary auditory cortex of primates. *Curr Biol, 13*, 554-562. Retrieved from http://download.current-biology.com/cellpress/pdfs/jcub/13/7/JCUB.13_7_554.2342.pdf

Wickelgren, B. G. (1971). Superior colliculus: some receptive field properties of bimodally responsive cells. *Science, 173*(3991), 69-72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4932262

Winer, J. A., & Schreiner, C. E. (Eds.). (2005). *The inferior colliculus*: Springer.

Woods, T. M., & Recanzone, G. H. (2004). Visually induced plasticity of auditory spatial perception in macaques. *Curr Biol, 14*(17), 1559-1564. doi:10.1016/j.cub.2004.08.059

Yamauchi, K., & Yamadori, T. (1982). Retinal projection to the inferior colliculus in the rat. *Acta Anat (Basel), 114*(4), 355-360. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6297221

Yasui, Y., Kayahara, T., Kuga, Y., & Nakano, K. (1990). Direct projections from the globus pallidus to the inferior colliculus in the rat. *Neurosci Lett, 115*(2-3), 121-125. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1700339

Yasui, Y., Nakano, K., Kayahara, T., & Mizuno, N. (1991). Non-dopaminergic projections from the substantia nigra pars lateralis to the inferior colliculus in the rat. *Brain Res, 559*(1), 139-144. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1723643

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat Rev Neurosci, 7*(6), 464-476. doi:10.1038/nrn1919

Yonovitz, A., & Harris, J. D. (1976). Eardrum displacement following stapedius muscle contraction. *Acta Otolaryngol, 81*(1-2), 1-15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1251700

Yonovitz, A., Lozar, J. T., Thompson, C., Ferrell, D. R., & Ross, M. (1977). "Fox-Box illusion": Simultaneous presentation of conflicting auditory and

visual CVs. *Journal of the Acoustical Society of America, 62*(S3 (Abstract)).

Zhang, A. B. (1984). Retinotectal pathways in rodents: particularly from the retinal ganglion cells to the inferior colliculus. *Taiwan Yi Xue Hui Za Zhi, 83*(1), 1-8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6586982

Zhang, M., Wang, X., & Goldberg, M. E. (2008). Monkey primary somatosensory cortex has a proprioceptive representation of eye position. *Prog Brain Res, 171*, 37-45. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18718280

Zhang, S. Q., Sun, X. D., & Jen, P. H. (1987). Anatomical study of neural projections to the superior colliculus of the big brown bat, Eptesicus fuscus. *Brain Res, 416*(2), 375-380. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3620966

Zhang, X., Guan, X., Nakmali, D., Palan, V., Pineda, M., & Gan, R. Z. (2014). Experimental and modeling study of human tympanic membrane motion in the

presence of middle ear liquid. *J Assoc Res Otolaryngol, 15*(6), 867-881. doi:10.1007/s10162-014-0482-8

Zhao, W., & Dhar, S. (2010). The effect of contralateral acoustic stimulation on spontaneous otoacoustic emissions. *J Assoc Res Otolaryngol, 11*(1), 53-67. doi:10.1007/s10162-009-0189-4

Zurita, P., Villa, A. E., de Ribaupierre, Y., de Ribaupierre, F., & Rouiller, E. M. (1994). Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. *Neurosci Res, 19*(3), 303-316. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8058206

Zwiers, M. P., Versnel, H., & Van Opstal, A. J. (2004). Involvement of monkey inferior colliculus in spatial hearing. *J Neurosci, 24*(17), 4145-4156. doi:10.1523/JNEUROSCI.0199-04.2004

Biography

Kurtis Gruters was born (November 3, 1985) and raised in Flagstaff, Arizona. He completed his undergraduate schooling in Rochester, New York where he received his Bachelor of Music degree in Music Theory with an emphasis in percussion performance from the Eastman School of Music and Bachelor of Science degree in Cognitive Neuroscience from the University of Rochester along with a Minor in Linguistics. His interest in music and language, along with his music performance experience, prompted him to attend graduate school at Duke University to study audiovisual interactions with his mentor, Jennifer Groh.

At Duke, Kurtis published a comprehensive review of non-auditory interactions in the inferior colliculus. He also completed the Certificate for Undergraduate Teaching program, was a Fellow in the Wireless Intelligent Sensor Network (WISeNet) program, and was a Cadet in the United States Army Reserve Officer Training Corps where he served as both Operations Officer and Battalion Commander. He completed his ROTC training as a Distinguished Military Graduate and was commissioned into the United States Army at the rank of Second Lieutenant in May, 2015.