J Speech Lang Hear Res. Author manuscript; available in PMC 2014 October 01.

Published in final edited form as:
J Speech Lang Hear Res. 2013 October ; 56(5): . doi:10.1044/1092-4388(2013/12-0283).

Auditory and Cognitive Effects of Aging on Perception of Environmental Sounds in Natural Auditory Scenes

Brian Gygi and
Veterans Affairs Northern California Health Care System, Martinez, CA; Nottingham Hearing Biomedical Research Unit, Nottingham UK

Valeriy Shafiro
Rush University Medical Center, Chicago, IL

Abstract

Purpose—When environmental sounds are semantically incongruent with the background scene (e.g., horse galloping in a restaurant) they can be identified more accurately by young normal hearing listeners (YNH) than sounds congruent with the scene (e.g., horse galloping at a racetrack). This study investigated how age and high frequency audibility affect this Incongruency Advantage (IA) effect.

Methods—In Experiments 1a and 1b, elderly listeners (N=18 for 1a, N=10 for 1b) with age-appropriate hearing (EAH) were tested on target sounds and auditory scenes in five Sound-to-Scene ratios (So/Sc) between −3 to −18 dB. Experiment 2 tested eleven YNH on the same sound/scene pairings lowpass-filtered at 4 kHz (YNH-4k).

Results—The EAH and YNH-4k groups exhibited an almost identical pattern of significant IAs, but both were at approximately 3.9 dB higher So/Sc than the previously tested YNH. However, the psychometric functions revealed a shallower slope for EAH compared to YNH for the Congruent stimuli only, suggesting a greater difficulty for the elderly in attending to sounds expected to occur in a scene.

Conclusions—These findings indicate that semantic relationship between environmental sounds in soundscapes are mediated by both audibility and cognitive factors and suggest a method for dissociating the two.

Introduction

Older listeners face ever-increasing challenges operating in the modern world of dense acoustic environments, such as streets, restaurants, offices and hospitals, all of which contain a multitude of objects and events competing for attention. A person's well-being can be directly affected by the inability to hear, recognize, and/or identify important sounds such as sirens, cell phone ringing and cars approaching. Although most normal hearing listeners can easily recognize an enormous number of these sounds in quiet (Gygi & Shafiro, 2010; Lass, Eastman, Parrish, K.A., & Ralph, 1982; Lawrence & Banks, 1973; Shafiro & Gygi, 2004), in everyday life multiple sound sources often co-occur, leading to masking of acoustic cues which are important for sound identification. This can have an especially deleterious effect on elderly listeners, who often have a high frequency hearing loss.

Correspondence concerning this article should be address to Brian Gygi, NIHR National Biomedical Research Unit in Hearing, 113 The Ropewalk, Nottingham UK NG1 5DU.

Numerous studies have extensively researched the effects of complex naturalistic backgrounds, usually mixtures of multiple talkers, on the perception of speech (often referred to as the "Cocktail party effect", or CPE, Cherry, 1953) in older as well as younger listeners (for reviews see Aydelott, Leech, & Crinion, 2010; Humes, 2008; Schneider et al., 2010). However, little is presently known about the effect of similarly complex auditory scenes (often referred to as soundscapes) on the perception of environmental sounds, despite everyday practical concerns and potential theoretical implications of environmental sound perception.

Speech comprehension in noise in the elderly is negatively affected by both peripheral auditory and central cognitive factors; however, the severe decline in auditory peripheral sensitivity found in nearly all elderly appears to be a major driver of these difficulties. In addition to a high frequency hearing loss and deficits in perception of temporal fine structure (Sheft, Shafiro, Lorenzi, McMullen, & Farrell, 2012), elderly have greater difficulties focusing on a single stream of speech in multi-talker situations and a diminished benefit from spatial localization (Dobreva, O'Neill, & Paige). Many cognitive abilities also suffer with aging, such as working memory span (Gilbert, 1941; Hasher & Zacks, 1988), speed of processing (Pichora-Fuller, 2003; Salthouse, 1996) and ability to inhibit task-irrelevant stimuli (Hasher, Zacks, & Rahhal, 1999; Kramer, Humphrey, Larish, Logan, & Strayer, 1994). However, some cognitive facilities, such as long-term memory and ability to apply prior knowledge seem to remain relatively undiminished with aging (Gilbert, 1941). Efforts to assess the relative effects of these factors have generally shown loss of audibility as the more prominent, accounting for about half the variance in speech processing (Divenyi & Haupt, 1997; Humes, et al., 1994). On many speech comprehension tasks such as the Speech-in-Noise test (SPIN; Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984), the elderly perform similarly to young when about 3 dB is added to the signal level (Schneider, Daneman, Murphy, & See, 2000). However cognitive abilities also play an important role, as shown by difficulties that persist even when audibility is corrected (e.g., with hearing aids) or controlled for (Humes, Lee, & Coughlin, 2006). The effect of cognitive abilities seems most apparent in elderly listeners when the task is complex and in demanding listening situations (Singh, Pichora-Fuller, & Schneider, 2010).

Although the major focus of most hearing research has been on speech, the ability to recognize environmental sounds is an important facet of hearing, playing a large role in listeners' safety and enjoyment (Andringa & Grootel, 2007; Ballas & Howard, 1987). As with speech, in real-world situations environmental sounds occur in contexts of other sounds. Although the perception of sounds in contexts is an important one, few have addressed this question for environmental sounds (Ballas & Mullins, 1991; Gygi & Shafiro, 2011; Leech, Gygi, Aydelott, & Dick, 2009; Niessen, Van Maanen, & Andringa, 2008) and fewer still have investigated how environmental sound perception changes with aging (Aydelott, et al., 2010).

Environmental sounds co-occurring in a natural scene may produce energetic masking when perceptually salient cues overlap in time and frequency and/or informational masking when individual sounds in a scene compete for attentional resources (e.g., a sound of car crashing may obscure the perception of water splashing even when both are presented at the same intensity level). In addition, semantic aspects of individual sounds in a scene may also influence perception. The nature of these effects, however, may not always be anticipated based on existing knowledge of context effects in speech perception. In the case of speech, congruent context, whether acoustic-phonetic, grammatical or semantic, leads to greater perceptual accuracy for individual words in sentences (which is the basis for the Speech-in-Noise test). Experimental evidence suggests that in speech perception the elderly rely on context even more than younger listeners to compensate for any losses in perceptual acuity

(Pichora-Fuller, Schneider, & Daneman, 1995). For example, perception of individual words or images can be facilitated by both aural and visual presentation of phonologically or semantically related primes; this effect is stronger in the elderly than in the young (Friederici, Steinhauer, & Frisch, 1999; Ganis and Kutas, 2003), just as perception of a visual object can be facilitated by the semantically congruent visual scene (Palmer, 1975). However, in the case of environmental sounds, the effects of context have not been well-studied. Although, based on speech and language research, one could expect a salubrious effect for sounds that are semantically congruent with their backgrounds (e.g. the sound of a cow mooing in a barnyard as opposed to a cow mooing in a bowling alley), as discussed in Gygi & Shafiro (2011), the contextual benefits of speech are quite different from that of natural auditory scenes, since grammatical context is far more constraining, and contextual effects of naturalistic backgrounds are largely probabilistic. As a result, to this point empirical findings regarding the effect of naturalistic backgrounds have been equivocal.

In one of the first published studies testing environmental sounds in context, Ballas and Mullins (1991) asked listeners to identify target environmental sounds embedded in sequences of other sounds which were either semantically congruent or incongruent. The researchers found that congruent context significantly increased performance above that of incongruent or neutral context; however, congruent context performance was not better than that obtained for the sounds presented in isolation without any context. More recently, Niessen, et al. (2008), using a similar paradigm, found that congruent context did help disambiguate sounds that were otherwise easily confused in quiet (what Ballas called "homonymous sounds"). However, the target environmental sounds in that study were digitally edited to contain conflicting acoustic cues for two different sounds and were likely to be more ambiguous than unedited environmental sounds used by Ballas & Mullins.

In fact, a contrary effect was found in two other studies that tested the identifiability of unmodified target environmental sounds mixed in naturalistic auditory scene in normal hearing young adults (Gygi & Shafiro, 2011; Leech, et al., 2009). Environmental sounds that were semantically incongruent with the scene were actually better identified than sounds which were congruent with the scene, producing an incongruency advantage (IA) (termed the "pop-out effect" in Leech, et al.). The IA, which has been found across subjects and experimental conditions (Gygi & Shafiro, 2011; Leech, et al., 2009), ranged from 2–5% and corresponded to an approximately 2.2 dB decrease in signal level (i.e., sounds in incongruent scenes required ~ 2 dB less energy than sounds in congruent scenes to be equally identifiable).

Thus far the causes of the IA are not fully understood since only a few studies have addressed it. Since the listeners in both Leech, et al. and Gygi & Shafiro had normal hearing, audibility factors were minimized – in Gygi & Shafiro, the baseline identifiability of the target sounds in quiet accounted for very little variance in the findings. However, the fact that the same target sounds and auditory scenes were used in both congruent and incongruent stimulus pairs also suggests that the major difference between the two was largely due to cognitive factors such as semantic association.

The discrepancy between previous studies that found no IA-type effect and those that did may be explained by differences among the experimental paradigms. Whereas Ballas and Mullins (1991) and Niessen et al. (2008) presented sequences of isolated sounds, thus avoiding any energetic masking, both Gygi & Shafiro and Leech, et al. mixed target environmental sounds into actual recordings of natural auditory scenes which were either semantically congruent or incongruent with the target sound. Further, the IA effect appears to be level dependent, varying as a function of Sound-to-Scene energy ratio (So/Sc)[1], as shown by Gygi & Shafiro. At high So/Sc ratios there was a consistent IA, but at lower So/Sc

NIH-PA Author Manuscript

NIH-PA Author Manuscript

there was either no significant effect or a slight advantage for congruent sounds as illustrated by the psychometric functions for congruent and incongruent sounds in that work.

Gygi & Shafiro (2011) moreover found that the IA effect was affected by listener experience with the target sounds and auditory scenes. An overall IA of comparable magnitude was demonstrated both in YNH (experienced listeners who had extensive training in the identification of the corpus of sounds and scenes in isolation), and in YNH-N (naïve listeners who had no prior experience with the stimuli). However, for the IA was found at different So/Sc ratios for the YNH and YNH-N listeners: the psychometric functions for the naïve listeners were shifted to the right by about 4.5 dB, indicating that familiarity with the sounds changed the overall difficulty of the task, but did not necessarily change the relative identifiability of the congruent and incongruent sounds. Thus the manifestation of IA could be influenced not only by the audibility of the target sound in an auditory scene, but also by memory representations for the sounds and perceptual processing. This is not surprising since by definition the congruence or incongruence of a sound with a scene is a semantic feature; as such it can also be expected to be influenced by cognitive factors.

Recently, Aydelott, et al. (2010) tested the same target sounds and auditory scenes used in Leech, et al. (2009) with elderly listeners. They found no difference between the congruent and incongruent conditions and interpreted this as demonstrating that "the older participants are unable to build up meaningful interpretation of the background auditory scene and so are unable to take advantage of this in detection of incongruent targets." A major difference between the paradigms in both Aydelott, et al., and Leech, et al., compared to Gygi & Shafiro, was that in the latter the participants were both familiarized with the auditory scenes to be used and informed via a text label what scene they were listening to, in order to make sure the context for the scene was sufficiently established. In contrast, in Aydelott et al. and in Leech et al., the participants had to first determine the scene they were in, and then identify the target sound. This is quite a challenging task for elderly listeners, due both to audiometric and cognitive factors, which may have led to the inability build up a meaningful interpretation of the scene. In addition, Aydelott et al. only tested sounds at fairly high So/Sc ratios: +3 and −6 dB. A breakdown of performance by So/Sc was not provided, but overall performance was extremely good, ranging from p(c) = 0.9 to 0.96 for the elderly. Thus the elderly were performing near ceiling levels where IA would not be measurable. The fact that the elderly listeners' puretone thresholds did not correlate with performance supports this interpretation; the sounds were audible enough that peripheral energetic masking did not play a role, and since they were not informed of the scene, it might be that contextual priming from the background scene also did not have a significant effect. Thus, the effect of context on perception of environmental sounds for elderly listeners in challenging listening conditions remains to be elucidated.

The present experiments investigated the effect of age and high frequency hearing loss on the identification of environmental sounds embedded in naturalistic auditory scenes, which were either congruent or incongruent with the target sounds, at various So/Sc levels. In Experiment 1, the sounds and scenes from Gygi & Shafiro (2011) were presented to elderly listeners with age-appropriate hearing ability (EAH – defined in the Methods), first at the So/Sc range used in the previous study (−18, −15 and −12 dB), and then at higher So/Sc. It was expected that the experienced older listeners (i.e., ones that had been familiarized with the sounds and scenes) would be less accurate overall than the experienced young normal-hearing (YNH) listeners tested by Gygi & Shafiro (2011) due largely to reduced audibility in

[1]The Sound to Scene ratio (So/Sc), defined in Gygi and Shafiro (2011), is the ratio of the pause-corrected energy of the target sound (rms minus pauses of more than 50 ms) to the pause-corrected energy of the segment of the background scene in which the target appears. Since it is different from traditional Signal-to-Noise (SNR) measures we do not use the standard nomenclature.

high frequencies. It was anticipated that the IA in EAH listeners would be found at higher So/Sc than YNH listeners. Furthermore, as was the case with naïve young normal-hearing listeners (YNH-N) in the previous study by Gygi & Shafiro (who were tested without being familiarized with the sounds), the psychometric function for the EAH listeners would likely be shifted to the right relative to the experienced YNH listeners. The magnitude of that shift would indicate the influence of aging and hearing loss on the perception of environmental sounds in natural auditory scenes.

In Experiment 2, to separate the effects of high frequency audibility from those of cognition, a new group of young normal hearing listeners were tested on the same corpus of sounds and scenes which were lowpass filtered at 4kHz (termed the YNH-4k listeners) to mimic the effects of reduced high frequency audibility in the elderly listeners from Experiment 1 resulting from hearing loss. It was expected that overall identification accuracy with the filtered stimuli would be lower than for the full spectrum stimuli in YNH listeners. If indeed audibility were the primary determinant of the IA, there would be no significant differences between the YNH-4k and EAH from Experiment 1 in terms of overall accuracy or psychometric functions. The findings from these experiments will illuminate the intertwined roles of audiometric and cognitive factors in the complex tasks of listening to a specific sound in a dense auditory scene.

Methods

The experiments described below tested the identification of environmental sounds mixed with familiar, naturally-occurring auditory scenes. The target sounds were either contextually “congruent”, i.e., likely to occur in a particular background scene (such as hammering at a construction site) or contextually “incongruent” (such as a horse galloping in a restaurant). Experiments 1a and 1b tested older listeners with age-appropriate hearing as defined above on identification of environmental sounds occurring in either congruent or incongruent auditory scenes across different So/Sc energy ratios. Experiment 2 tested young normal hearing listeners on identification of the same sound-scene pairs, which were low-pass filtered to simulate the general effect of high frequency hearing loss. For consistency with previous work with YNH listeners, the stimuli and procedures for both experiments closely followed those used in Experiment 3 of Gygi & Shafiro (2011).

Stimuli

The target sounds to be identified were thirty-one familiar environmental sounds selected from previous studies (e.g., Gygi, Kidd, & Watson, 2004, 2007; Shafiro, Gygi, Cheng, Mulvey, & Holmes, 2008). These sounds represented three different major sound source categories derived in Gygi, et al. (2007): harmonic sounds, continuous non-harmonic sounds and impact sounds. To increase stimulus variability, each target environmental sound was represented by 3 – 5 different tokens, e.g., three different Match Striking sounds, or five different Baby Crying tokens. The tokens for each sound were selected to be as acoustically different as possible, while still being representative of the sound type. All target sound tokens were obtained from high-quality sound effects CDs (Hollywood Leading Edge and Sound FX The General) sampled at 44.1 kHz. The target sounds ranged in length from 457 ms (Sneeze) to 3561 ms (Baby Crying).

All target sounds were equated for overall root mean square energy (rms), with a correction for pauses of greater than 50 ms (see Gygi, et al., 2004 for a description of the pause-corrected rms). In baseline identifiability studies with young normal-hearing listeners (Gygi & Shafiro, 2011), all tokens were found to be highly identifiable and typical representations of target sounds. Auditory scenes used as backgrounds consisted of fourteen high-quality recordings of common and highly identifiable natural auditory scenes, donated by professional sound recordists. The recordings were all in stereo and recorded at a sampling rate of 44.1 kHz or higher (if higher, they were downsampled to 44.1 kHz). All scenes were edited to be ten seconds long, and similar to the target sounds, were equated for overall rms, with a correction for pauses of greater than 50 ms. The scenes represented a variety of indoor/outdoor and urban/rural settings. In rare instances when intelligible human speech could be heard in a scene, it was not specific to the context of the scene and was not expected to help the listener in identification of the scene. In pilot studies testing identifiability of the scenes themselves with young normal-hearing listeners, twelve of the fourteen scenes were identified at p(c) (percent correct) > 0.8.

The sound/scene pairs, listed in detail in Table 1 of Gygi & Shafiro (2011), were designed to be either congruent or incongruent, or foils, which are described separately below. (In)congruency between the target sounds and the auditory scenes was initially determined by three judges who were familiar with the study (including the first author) and later confirmed in a separate laboratory-based study and also in another Web-based congruency rating study, as reported in Gygi & Shafiro (2011). Each scene was paired with two congruent and incongruent target sounds, for a total of 56 sound-scene combinations. Consequently, twenty-five of the target sounds were presented in both congruent and incongruent situations. However, six sounds were only presented in either a congruent or incongruent scene because a complete counterbalancing of sounds and scenes was not possible for all sounds due to some sounds occurring only in particular congruent scenes. Each target sound occurred in an auditory scene only once. For example, if the target sound was a cow mooing and the scene was a barnyard scene, there would be no other cow mooing in the barnyard scene. Overall, taking into account the different congruent/incongruent pairings and the various tokens for each target sound, there were a total of 217 congruent and incongruent sound-scene combinations. Some examples of the sound-scene pairs used as stimuli are included in the Supplemental Materials.

In addition to congruent and incongruent pairings, 36 sound-scene combinations were used as foils. The foil stimuli were constructed from six of the target sounds and fourteen of the scenes used for the congruent and incongruent pairings as well as fifteen novel target sounds and four novel scenes. The foils were constructed to be neither clearly congruent, nor incongruent. Their purpose was to mitigate any learning by participants of the specific sound-scene pairs, by increasing uncertainty as to the probable response alternatives; thus, some of the foils comprised sounds and scenes that were also targets, so listeners learned they could be used in multiple pairings. The foils were not selected to be neutral with regard to context; rather, they were chosen without taking into consideration their congruence or incongruence. Including foils, there were 253 sound-scene combinations used at each So/Sc.

The target sounds began four seconds into the scene. The onsets and offsets of the 3.5 sec listening interval containing the target sound were marked with “ding” sounds. There was always 100 ms between the first ding and the onset of the target sound, as well as at least 100 ms between the offset of the target sound and the final ding. Although these acoustic markers were quite salient and might have the effect of distracting the listener, pilot studies for Gygi & Shafiro (2011) suggested that, given the length of the auditory scenes and variety of target sound sources, there was a much greater amount of uncertainty among the listeners as to what the target may be when a more specific listening interval in the scene was not marked.

The stimuli were mixed at different Sound-to-Scene ratios (So/Sc), which were determined by comparing the pause-corrected rms of the target sounds to the pause-corrected rms of the section of the scene in which the target sounds appeared. In Experiment 1a and Experiment 2, low signal to noise ratios were used. All tokens for each of the target sounds were mixed with the auditory scenes in the combinations described above at So/Sc of −18, −15 and −12 dB, to match the levels used in testing young normal hearing listeners in Experiment 3 of Gygi & Shafiro (2011). In Experiment 1b, the stimuli were similar to those in Experiment 1a, except the target sounds were mixed with the auditory scenes at higher So/Sc of −3, −6, and −9 dB.

Procedure

In all the experiments described here, the trial procedure was the same. Seated in a soundproof booth in front of a computer terminal and keyboard, listeners were told about the nature of the study. Specifically, that this study was to determine how well they could identify environmental sounds in naturally occurring scenes. During the initial briefing listeners were told that "some sounds may seem perfectly natural, but some sounds might be a little unusual in a particular scene," but they were not told of the focus on the difference between congruent and incongruent sounds. They were given a sheet listing the sounds to be presented and three-letter codes with which to respond. There were 90 possible codes with code labels intended to give a sense of the source objects and the events involved (e.g., CAT for cat meowing, SPL for splash in water). The total number of codes was greater than the number of target sounds to approximate open-set conditions of everyday listening and minimize closed-set effects, while still ensuring standardization of responses. This same three-letter code format was used successfully for identification in Gygi, et al. (2004) and Gygi & Shafiro (2011). All listeners were then given two practice trials with environmental sounds, which were presented at very high levels.

After the practice trials they the participants were given a response code familiarization session, in which each possible target sound was played in quiet. Listeners were presented with a label describing a sound and told to type in the corresponding code. This familiarization reduced the time listeners spent searching for the appropriate key code. The elderly listeners were asked if they had difficulty learning the key codes, and none reported any problems. In addition, the elderly did not take substantially longer to complete a block of trials than the young listeners – both groups could complete a block in about an hour. So it appears the use of key codes did not unduly burden the elderly listeners.

On every trial one sound/scene mixture was presented diotically through headphones, while a label for the scene the listener was hearing was displayed (e.g., "Street at Night") throughout the presentation of the scene. The scene label was shown to avoid any confusion about the scene they were listening to (which as Aydelott et al., 2010, showed, can strongly influence the results), so that he/she could focus on the cues to identifying the target rather than the scene. This would minimize differences that may arise due to variation in scene baseline identifiability. To further mark the listening interval, during the presentation of an auditory scene, at the onset of the listening interval that contained the target sound, the text box changed color from yellow to orange and remained that way until the offset of the interval, thus informing the listener of when to listen for the target sound. Listeners were instructed to identify the target sound by typing the appropriate code on the keyboard. The list of codes was always within view in front of each subject. If the listener did not respond within seven seconds, a prompt would flash on the screen, encouraging them to respond. If the listener responded with a code that was not valid, they were prompted to reenter the code. No feedback was provided. All responses were recorded electronically and saved on a file server.

In each experiment reported below the stimuli were blocked by So/Sc and randomized within each block. The blocks were presented in an ascending order of So/Sc with the lowest So/Sc stimulus block always presented first, to minimize any learning effects resulting from hearing the target sounds at easier So/Sc. Listeners were given a break after completion of a particular So/Sc level, and told that on the next block, the targets should be easier to identify. Stimuli were generated from digital files by Echo Gina 24 sound cards, amplified by the TDT System 2 headphone buffer and presented through high-quality Sennheiser 250 II headphones. Prior to testing, a 1-kHz calibration tone of the same rms as the equated level of the auditory scenes was set to 75 dB SPL at the headphones.

In both experiments, after the code familiarization session, three additional familiarization sessions with sounds and scenes were conducted. Familiarization sessions consisted of the following: sound identification in quiet, sound typicality ratings and scene identification. For sound identification in quiet, listeners were tested in quiet on baseline identification of 195 environmental sounds, including multiple tokens of 51 unique sounds and including all sounds to be used as targets. The listeners identified each sound in quiet, presented separately, using the codes described above. No feedback was given on these baseline trials. In the next testing session the listeners rated the typicality of each of the tokens on a scale of 1 to 7. The typicality ratings were only used to validate the particular tokens selected as targets (the data for young listeners are included in Gygi & Shafiro, 2011). In the final session the listeners performed open-set identification of the auditory scenes (the listeners described the scenes in their own words). No feedback was given during these familiarization sessions. Each session required one hour to complete and were run on three separate days. Overall, after the listeners performed identification of isolated sounds and scenes in quiet, they were highly familiar and practiced with both the target sounds to be identified and the auditory scenes the sounds were embedded in before beginning the experiment described below.

Experiment 1a – Elderly Listeners at Low So/Sc

Subjects

Nineteen elderly listeners (10 males and 9 females) between the ages of 60–80, were tested at the Speech and Hearing Research Laboratory at the Veterans Affairs Northern California Health Care System Martinez Outpatient Clinic in Martinez, CA. All except one were native English speakers, were in good health, had normal (> 25) scores on the Mini-Mental State Test (Folstein, Folstein, & McHugh, 1975) and had no history of mental or otological disorders. They all had pure-tone thresholds of 40 dB HL (ANSI 1969 Hearing Level) or better at 500, 1000, 2000, and 4000 Hz, and 70 dB HL or better at 6 and 8 kHz. Although they all had some hearing impairment at higher frequencies, based on the ISO 7029 (2000) statistical distribution of hearing thresholds as a function of age, the largest mean deviation from the median threshold across frequencies for any of them was +1.12 *SD* (i.e., the worst listener had thresholds 1.12 *SD* higher than the median). Thus, for this study they are designated Elderly with Age-appropriate Hearing (EAH) listeners. The pure tone thresholds for the EAH participants are presented in Figure 1. Subjects were recruited by flyers and paid for their participation. One of the listeners had very poor performance on the testing described here and this subject's data were not included in the analysis of results (it was later found out that he was not actually a native English speaker, although he had lived in the US for fifty years). Approval for all the protocols described herein were obtained from the U.S. Department of Veterans Affairs Northern California Health Care System Institutional Review Board and all subjects were handled in accordance with these protocols.

Results and discussion

Percent correct accuracy scores for each listener in each So/Sc and congruency condition were converted to rationalized arcsine units (RAU) which normalizes the variance for p(c) data (Studebaker, 1985), and subjected to a repeated measures 3×2 ANOVA, with So/Sc as one factor and Congruency as another. As in Gygi & Shafiro (2011) performance improved with higher So/Sc, leading to a main effect of So/Sc, $F(2, 34)=339.26$, $p<.001$, but no main effect of Congruency was detected. However, there was an interaction of So/Sc and Congruency $F(4, 68) = 8.24$, $p<.001$, which was likely due to better performance in the Congruent condition at the lowest So/Sc, −18 dB, by almost all the listeners. A posthoc LSD test showed the Congruent and Incongruent conditions at that So/Sc to be significantly different, $p < .001$, with an effect size of Cohen's $d = 1.577$, which is a large effect as described in Cohen (1988), whereas at −15 dB it was approaching significance, $p = 0.088$. In general, there was large intersubject variability, as illustrated in Figure 2, which shows the results for each listener by So/Sc and Congruency condition. In Figure 2 only the Incongruent and Congruent conditions, which are the focus of this analysis, are included for clarity. The data from Experiments 1a and 1b are presented together, so the data from Experiment 1a occupy the right three panels of Figure 2. In most cases the foil results were between the Incongruent and Congruent results; there was no instance in which the foils' p(c) was significantly greater or less than for the two main groups. Despite the range in individual performance, the pattern of results was remarkably similar across participants. Most intersubject correlations were significant and the mean intersubject correlation was 0.82. Nevertheless, there was a demarcation in performance in Congruent and Incongruent conditions at the two lower So/Sc (−15 and −18 dB) based on overall performance. It appears that the lower-performing subjects actually had a greater difference between the Congruent and Incongruent data points at each So/Sc than the better-performing ones. The EAH subjects were divided into two subgroups based on their p(c) in the Incongruent condition at the −12 dB So/Sc. On Figure 2 the subgroups are marked by two ovals, which resulted in an N=8 in the "Higher performing" group and N=10 in the "Lower performing" group. To more easily distinguish the two subgroups, the data points for the Lower-performing are all represented by filled symbols. A more detailed discussion of the relative performance of the Higher- and Lower-performing subgroups follows below in conjunction with Figure 4.

Next, the mean performance of the elderly listeners in this experiment was compared with that of the experienced young normal-hearing listeners (YNH) tested in Experiment 3 of Gygi & Shafiro (2011), who had received the same level of pretest exposure to the target sounds and auditory scenes. Figure 3 shows the performance of both groups in the same conditions. It can be clearly seen that despite similar amount of exposure prior to the test, the elderly performed worse overall (i.e., lower p(c)) than the YNH listeners at every So/Sc. A repeated-measures between-groups ANOVA showed a main effect of age group, $F(1, 26) = 14.96$, $p < 0.001$, and a main effect of So/Sc, $F(2, 52) = 224.19$, $p < 0.001$ (both groups got better with higher So/Sc). However, there was no main effect of Congruency, but there was a significant interaction of age group and Congruency, $F(1, 26) = 10.50$, $p = 0.0033$. There was also an interaction of So/Sc and Congruency, $F(2, 52) = 18.43$, $p < .001$, since the effects of Congruency varied with the So/Sc for both groups, but there were no interactions of So/Sc with age group. Moreover, it was not the case that the young and elderly were preferentially attending to different sounds. The correlations of the p(c) for the specific sound tokens in each So/Sc for the two groups ranged from $r = 0.53$ at −18 dB (at which many tokens were not identified by the elderly at all) to 0.86 at −12 dB, indicating that both groups of listeners performed well and poorly on the same target sounds.

In addition to the lower overall p(c), EAH listeners differed from that of YNH listeners tested by Gygi & Shafiro (2011) in that the EAH did not show an IA at −12 and −15 dB So/Sc, unlike the YNH in the earlier study who had a significant IA at these So/Sc levels. Furthermore, the EAH actually performed better in the Congruent condition at the lowest So/Sc, whereas the young listeners showed no difference. That performance was better for Congruent sounds at the lowest So/Sc can be explained by guessing strategies. As shown in the Gygi and Shafiro 2011 study, when listeners were wrong, they had a strong tendency to select sounds that were congruent to the scene that was being presented. This suggested that when listeners were not sure of the target sound, they guessed sounds that would normally occur in the relevant scene. This would tend to increase the p(c) for congruent sounds and lower it for incongruent[2]. Notably, posthoc comparisons show no difference in the p(c) for the YNH at – 15dB So/Sc from that of the EAH at −12 dB So/Sc, suggesting that the entire psychometric function for the EAH might have been shifted relative to the YNH by > 3 dB, which is similar to what has been found to equate performance in speech comprehension tasks (Singh, et al., 2010).

As mentioned above, there were two different patterns of performance at the lower So/Sc (−15 and −18 dB So/Sc) between the higher- and lower-performing participants, which are delineated in Figure 2 by two ovals (resulting in an N=8 for the higher-performing subgroup and N=10 for the lower-performing subgroup). The mean p(c) of the two subgroups are plotted separately in Figure 4. An ANOVA on the two subgroups was performed and posthoc Fisher LSD tests indicated that the lower-performing subgroup actually showed a significant advantage for *Congruent* sounds at the lower So/Sc, −18 and −15 dB. Possible reasons for this Congruency Advantage (CA), when individual environmental sounds congruent with the background scene are identified with a greater accuracy (also found in YNH in Gygi and Shafiro, 2011), are presented in the General Discussion. In contrast, the higher-performing participants showed an Incongruency Advantage at −12 dB, which is what was found with the YNH listeners in Gygi & Shafiro (2011).

Notably, performance of the older listeners in the present experiment was similar to that of a second group of listeners, young normal-hearing naive listeners (YNH-N) tested in Gygi & Shafiro (2011). These participants did not receive baseline training on the target sounds nor participated in identification of the auditory scenes prior to testing. YNH-N did not show an IA at the same conditions as the YNH; for them IA was manifest only at – 9 dB So/Sc or higher, and their psychometric functions were shifted by 4.5 to higher So/Sc compared to those of the experienced YNH listeners. So the effects of aging and the lack of familiarity with the stimuli appear to produce comparable results in older and younger subjects, respectively, i.e. to worsen performance at low So/Sc for both Incongruent and Congruent sounds.

Overall, the present findings are consistent with those from Gygi & Shafiro (2011) and indicate that the IA is a So/Sc-dependent effect that could be found above the midpoint of the psychometric function. Moreover, the pattern of results for the elderly experienced listeners (EAH) seem to follow that for YNH-N (young naïve listeners) tested earlier in that both groups were different from the YNH in similar ways. The YNH-N listeners exhibited a comparable IA to the YNH listeners, but at a higher So/Sc (in that case ~4.5 dB higher So/Sc). Therefore, it could then be expected that EAH listeners, like YNH-N listeners, will demonstrate IA at more favorable So/Sc ratios than those tested in the present experiment. This hypothesis was tested in Experiment 1b.

Experiment 1b – Elderly Listeners at Higher So/Sc

Procedure

Unlike Experiment 1a, the subjects were only tested on identification of sounds in scenes because they had all previously undergone the familiarization with the sounds and scenes used as stimuli. The participants were informed that this test would be at higher So/Sc levels

[2]Although the results by pure guessing would be 1/90 = .0111, not all the sounds have the same *a priori* probabilities in a given scene. What exactly those are is not known, but even assuming listeners had ascertained the four target sounds for each scene that would put their level of chance at 0.25.

and thus would likely be easier than the last one. As in Experiment 1a, all the stimuli at the lowest So/Sc were presented first in random order, then the stimuli at the −6 dB So/Sc in a new random order, and finally the stimuli at the −3 dB So/Sc again in a different random order.

Subjects

Ten of the EAH listeners from Experiment 1a (4 males) were able to return to the laboratory approximately four months after Experiment 1a. Five of these participants had been among the higher-performing group in Experiment 1a, and five were from the lower-performing group. None had reported any changes in their hearing or cognition since the last test.

Results

As in Experiment 1a, individual accuracy scores were converted to RAUs and subjected to a 3 × 2 repeated measures ANOVA. There was a main effect of So/Sc, $F(2, 18) = 47.36$, $p < .001$, an interaction of So/Sc, $F(4, 36) = 3.1488$, $p = .0256$, but, as previously, no main effect of Congruency. The p(c) for each congruency condition in Experiment 1a and 1b are plotted in Figure 3 as a function of So/Sc. Performance again improved monotically with So/Sc, and there were significant IAs in both the −9 and −6 dB So/Sc conditions of p(c) = 0.086 and 0.029, respectively, which are comparable to the IAs found for experienced YNH listeners at lower So/Sc in Gygi & Shafiro (2011). The effect sizes as measured by Cohen's *d* were .898 at −9 So/Sc and .931 at −6 dB, both of which are considered medium sized effects (Cohen, 1988). At −3 dB So/Sc there was no significant difference between the congruency conditions, likely due to ceiling effects. As in Experiment 1a, the intersubject correlations were relatively high, with a mean of $r = 0.61$. Individual performance is displayed in the left three panels of Figure 2 across So/Sc and congruency conditions along with the results from Exp. 1a in the right three panels. Thus, as expected, it seems that the elderly do indeed exhibit an Incongruency Advantage, but at a higher So/Sc than the experienced young listeners.

To verify that the results of Experiment 1a are comparable with those of Experiment 1b despite the fact that some participants from Experiment 1a did not participate in Experiment 1b, an ANOVA was conducted only on the data from the ten older listeners that took part in both Experiments 1a and 1b at all So/Sc ratios. Comparing the performance of the EAH listeners who participated in all So/Sc (n=10) compared to those who participated in only the low So/Sc (n=8) showed that no main effect of listener group was found, $F(1, 16) = 2.055$, $p = .171$, and the effects found in Experiment 1a for all 18 listeners were also found in the subset of 10 who took part in both Experiment 1a and 1b (with a single exception at −18 So/Sc). Thus, including data from all participants in the analysis seemed justified.

The psychometric functions for the EAH listeners were next plotted as a function of So/Sc, and were compared to the psychometric functions for the experienced YNH listeners from Gygi & Shafiro (2011), as shown in Figure 5, along with the values corresponding to 0.2, 0.5 and 0.8 p(c) isoperformance points. Logistic psychometric functions of the form

$$\psi(x;\alpha,\beta,\gamma,\lambda)=\gamma+(1-\gamma-\lambda)\,F(x,\alpha,\beta)$$

were fitted to the individual EAH data separately for each congruence condition using psignifit version 2.5.6 (see http://bootstrap-software.org/psignifit/), a software package which implements the maximum-likelihood method described by Wichmann & Hill (2001a). These were derived by the BCa bootstrap method implemented by psignifit, based on 4999 simulations (see Wichmann & Hill, 2001b). The resulting fits were quite good: the psychometric functions accounted for 97.3% and 96.5% of the variance in the Congruent and Incongruent data, respectively. In addition, the correlations between the model predictions and the deviance residuals (which might indicate systematic errors in the model, if significant) were quite low, r = 0.031 and 0.013 for the Congruent and Incongruent fits, respectively.

Figure 5 indicates that the overall functions for the EAH are shifted about 3.9 dB to the right of the YNH, from a mean of −17.6 dB (across Congruency condition) to −13.7, reflecting poorer overall performance. This corresponds with the ~3 dB extra energy needed to equate the performance of elderly and young listeners found in studies of speech perception in noise (Pichora-Fuller, et al., 1995). In general, the relative shapes of the psychometric functions are very similar for both groups, in that the slopes of the congruent functions are shallower than for the incongruent functions. Thus, both groups show an IA at higher So/Sc and an advantage for congruent sounds (CA) at lower So/Sc. However, there are also some differences in IA between the YNH and EAH listeners that can be seen in the psychometric functions. For the isoperformance point corresponding to p(c) = 0.8, the IA, that is, the dB difference in the congruent and incongruent functions at that p(c), is 2.2 dB for the YNH group (reported in Gygi & Shafiro, 2011) and 3.0 dB for the EAH group. This indicates that the EAH listeners might be able to benefit from incongruent context, compared to congruent context, more than the YNH listeners[3].

To examine whether this difference was significant, and whether it resulted from EAH listeners' better processing of sounds in incongruent contexts or worse processing in congruent context, a further cross-group comparison of the slopes of the psychometric functions was conducted, along with the YNH-N data from Gygi and Shafiro (2011). The results are illustrated in Figure 6 displayed by congruency condition. Statistical tests of significance for the differences in the slopes were calculated using psignift, described above, which performed Monte Carlo simulations to estimate the probability that densities from the Monte Carlo distribution are smaller than that from the observed difference. Although it revealed no significant differences in the slopes of the Incongruent functions across EAH, YNH and YNH-N (naïve) subjects, which were .0562, .0478 and .0624 respectively, the Congruent function for the EAH listeners had a considerably shallower slope in comparison to the YNH naïve group. This was confirmed by significance tests, showing the slope for the EAH (.0356) to be significantly less than for the YNH-N group (.0536), so much so that the two functions touch at high So/Sc. However, the differences in slope were not significant between EAH and YNH groups who had the same level of experience with the stimuli. The slopes and thresholds at the 0.80 isoperformance point for the psychometric functions for all the groups in both Gygi & Shafiro (2011) and this study are shown in Table 1, along with the 95% Confidence Intervals (CI) for the data presented in this study. The difference in slopes between EAH and YNH-N listeners may indicate that the greater IA for EAH at high So/Sc is actually due to their greater difficulty with sound identification in Congruent context rather than a greater facility with sounds in Incongruent context when effects of age and hearing loss are combined with the lack of familiarity with the stimuli. This is discussed more fully with relation to the results of Experiment 2 below.

One possible reason for the difference between young and older listeners may be largely due to audibility. Although the present EAH listeners had age-normed hearing levels, their audiometric thresholds, particularly at high frequencies, were considerably elevated

[3]Unfortunately the significance of the EAH IA veruses the YNH IA cannot be tested via parametric methods (since the distributions from the simulations cannot be assumed to be normal). The IA predicted by the model at 0.8 p(c) for the EAH was 3.0 dB. From the distribution of the simulations for the YNH results from Gygi & Shafiro (2011) an IA of that size is greater than 70.7% of the simulated IAs. Thus the expected IA for the EAH is larger than the majority of the IA for the YNH. While not a definite significance test, it does indicate that we could expect to find a larger IA for the EAH than the YNH in the majority of instances.

compared to the <15 dB HL range demonstrated by YNH listeners. As can be seen in Figure 1 most of the EAH listeners had a roll off from 4kHz to 8kHz, from a mean of ~25 dB HL to ~40 dB HL, although there was wide variation. The correlations of the mean difference of the individual thresholds (calculated as 8kHz minus 500 Hz thresholds) for each listener are shown in Figure 7. The threshold difference for the right ear correlated significantly with the p(c) in both congruency conditions at −12 and −15 dB, and the threshold difference of the left ear correlated significantly at −12 dB. Importantly, the relationship between the threshold differences and p(c) declined gradually with increasing So/Sc. This suggests that hearing acuity did affect performance, and more strongly so when the audibility of the target sound was greater. This again resembles findings of elderly listeners with speech stimuli that show that audiometric factors are more predictive in easier listening conditions (Singh et al., 2010), whereas cognitive factors become more relevant in more difficult listening conditions. On the other hand, in the current study other hearing acuity-based measures (pure tone averages between 0.5 – 4KHz, better ear thresholds, 8kHz minus 2kHz threshold differences), were poor predictors of performance, which might have been partly due to the low N (10). It is worth noting that although the higher-performing subgroup of the elderly listeners had lower thresholds than the lower-performing group on all the audiometric measures, none of them were significant, possibly due to the large variance and a relatively small number of subjects in each group.

To further investigate the effect of reduced audibility in high frequencies on IA and sound identification performance, in Experiment 2 a new group of young normal listeners was tested using the same procedures as before; but in this case the sounds were low-pass filtered at 4 kHz to simulate the loss of high frequency information experienced by the EAH listeners (named the YNH-4k listeners). If the appearance of the IA at higher So/Sc for the EAH than for the YNH listeners was primarily due to reduced audibility of higher frequency acoustic cues, then similar IA patterns should be present in this group of YNH-4k listeners as in the EAH listeners.

Experiment 2 – Young Normal Hearing Listeners with Lowpass Filtered Stimuli (YNH-4k)

Stimuli and Procedure

The stimuli were similar to those in Experiment 1, except that the stimuli (sound and scene) were low-pass filtered using 6th-order Butterworth filters with a 4 kHz cutoff. The Butterworth filters were designed and implemented in Matlab. As a result, these stimuli have a drop of half-power at 4 kHz and are 40 dB down at 8 kHz, which approximates the high frequency hearing loss found in the EAH participants in Experiment 1, described above. The procedure was the same as in Experiment 1: subjects were first familiarized with the target sounds, the auditory scenes, and the response codes to be used, before testing the mixed sounds and scenes. Thus, the present subjects had a comparable amount of experience with the sounds and scenes as the EAH listeners tested before. The So/Sc levels used were, as in Experiment 1a, −12, −15 and −18 dB So/Sc.

Subjects

Participants were eleven young, adult, normal-hearing subjects between the ages of 18 – 30 (five males) who had not taken part any of the previous testing. All had normal hearing as measured by pure tone audiograms (thresholds < 15 dB HL from 500 – 8000 Hz).

Results and Discussion

As shown in Figure 8, the young normal listeners tested with lowpass stimuli (YNH-4k) were similar to the EAH in that they exhibited an advantage for congruent sounds (CA) at the lowest So/Sc, −18 dB So/Sc, but they also showed an Incongruency Advantage at a higher So/Sc than the EAH (in this case, −15 dB). There was no significant difference between the Congruent and Incongruent conditions at −12 dB So/Sc, unlike the results for similarly experienced YNH from Gygi & Shafiro (2011). Comparing the means at each So/Sc/Congruency condition for the YNH-4k listeners with the EAH listeners in Experiment 1, there were no significant differences between the two groups at any So/Sc or Congruency condition. Thus the EAH performed similarly to the YNH-4k listeners tested with filtered stimuli; in fact, at both the −12 dB and −18 dB So/Sc, the groups performed almost identically. Although there was an apparent difference between the groups at −15 dB, it was not significant. However, this difference does affect psychometric functions for the two groups, as described in the General Discussion. Overall, the results of Experiment 2 suggest that high frequency audibility can influence the presentation of the IA effect across So/Sc energy ratios, and can make YNH listeners with normal hearing perform similarly to EAH.

General Discussion

The present findings demonstrate that at higher So/Sc EAH listeners identify environmental sounds in an incongruent auditory context more accurately than sounds presented in a congruent context. This finding is consistent with the IA effect previously shown in the YNH listeners in that both groups show an IA. However, the psychometric functions of the two groups suggest that the IA may actually be greater in magnitude for the older than for the young listeners with a similar amount of experience with the target sounds and auditory scenes (who heard the unfiltered stimuli).

The identifiability of sounds in both congruent and incongruent contexts is determined by a mix of both basic auditory (signal audibility and coding of details by the auditory periphery) and cognitive (contextual semantic relation of background scene and target sound) factors. The psychometric functions of different subject groups reflect this interaction in complex ways. The 0.5 p(c) thresholds seem to represent the overall difficulty of the task, and so can be affected adversely by either reductions in audibility or cognitive decline. In many auditory training tasks, the psychometric function tends to move to the left as the listeners become more facile at the task (Leek & Watson, 1984). This was the case in the previous study, Gygi & Shafiro (2011), where YNH listeners experienced with the target sounds and auditory scenes outperformed naïve YNH-N listeners who were not familiar with the sounds or scenes. However, the thresholds at the 0.5 p(c) point on the psychometric function for the YNH listeners were significantly better than for both the present YNH-4k participants who heard the 4kHz lowpass filtered stimuli and the EAH listeners, even though all groups had the same level of pre-test training with sounds and scenes. Both YNH-4k and EAH groups were about 3 dB worse overall than the YNH who heard the full spectrum stimuli, which is consistent with previous findings for the elderly using speech stimuli presented in noise (Pichora-Fuller, Schneider, & Daneman, 1995). Thus, this difference seems to be largely one of audibility.

In contrast to the 50% threshold point, the slopes of the psychometric functions may be affected by how well listeners can both detect and attend to salient features at different levels of difficulty, which is a more complex process. This is only partially a function of the energetic masking of the background auditory scenes as evidenced by a moderate correlation of threshold and slope, $r = 0.48$. That is, functions with higher (i.e., worse) thresholds tended to have steeper slopes and vice versa, a finding that has occasionally occurred in previous psychoacoustic research (Gold, Law, Connolly, & Bennur, 2010). This might account for the relatively shallower slopes of the experienced YNH for both the Congruent and Incongruent conditions (since the threshold for that group was much lower than for any other group). On the other hand, slope differences across the two congruency conditions may indicate the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

influence of non-energetic masking factors. Specifically, there were no significant differences in the 50% thresholds for the EAH as compared to the YNH-4k listeners; but the slopes for the EAH in the Congruent condition were shallower than for the YNH-4k as shown in Figure 9. This finding is not likely to be due to decreased high frequency audibility or differences in familiarity with the target sounds and auditory scenes, since those were equivalent for the two groups and the two congruency conditions. Instead, the shallower congruent slopes in the EAH listeners may reflect the influence of cognitive factors, such as selective attention, inhibition of irrelevant information and working memory, which are known to decline in the elderly (Li, Daneman, Qi, & Schneider, 2004; Meihong, et al., 2012; Schneider, et al., 2010).

The finding that the slopes were significantly shallower for the EAH only in the Congruent condition is also consistent with the conclusion of Gygi & Shafiro (2011) that participants may listen differently to environmental sounds that are congruent compared to those that are incongruent with an auditory scene. When attending to sounds and scenes as a whole, similar to what Helmholtz, (1875) termed the “synthetic” mode, or, more recently what Truax (2001) called “Listening in Readiness,” the elderly listeners seem to perform comparably to the young. In this way of listening, the various sources that make a scene are grouped together in a single auditory object, and thus it is easier to notice incongruent sounds that seem to “pop out” at the listener. In this case, rather than focusing on specific objects in a scene the listener perceives more global and holistic scene properties related to its scene structure and functional attributes. Recent research with visual scenes has provided empirical support for this approach (Greene and Oliva, 2009). The present data suggest that the elderly are just as good at young listeners at attending in this mode. However, for the Congruent sounds, since nothing stands out, the listeners need to switch to a more analytic mode of listening, or what Truax referred to as “Listening in Search.” In this case, the EAH listeners were at a notable disadvantage, which may reflect the strain of cognitive demands in examining the auditory image in working memory for the target sound. In general, the slopes seem to represent a complex mix of cognitive and peripheral factors including audibility, familiarity with the stimuli, ability to allocate attention and, possibly, degraded cochlear processing of audible sounds which characterizes presbacusis (Moore, 1998).

Summary and Future Directions

The results of these experiments show that the Incongruency Advantage is a complex effect that reflects both cognitive and auditory factors. The results demonstrate that older listeners obtain the Incongruency Advantage effect and thus are able to identify target environmental sounds with greater accuracy in naturalistic auditory scenes semantically incongruent with the sounds than in those congruent with them. This IA effect, however, is dependent on the So/Sc energy ratio – at higher So/Sc there is about 5–7 percentage point improvement, whereas at the lowest So/Sc there is actually an advantage to Congruent sounds. The IA further occurs at different levels for EAH vs. YNH listeners; the elderly require approximately 3.6 dB higher So/Sc to perform the same as the young listeners.

Much of that difference appears to result from loss of high frequency audibility for the target sounds and auditory scenes, as shown by the remarkably similar performance of the EAH and YNH-4k, young normal listeners presented with stimuli that were low pass filtered below 4 kHz (simulating the hearing loss in the elderly). This is consistent with the findings in speech research e.g., Schneider, Daneman, & Murphy (2005), suggesting that audibility is a major driver of performance in less difficult listening conditions (in this case, at higher So/Sc). The loss of high frequency audibility also leads to a decrease in overall performance accuracy for sound identification in naturalistic scenes.

On the other hand, as indicated by their shallower slopes of psychometric functions in congruent conditions, the elderly may experience greater difficulty than the young in identifying target environmental sounds in semantically congruent auditory scenes. The effect of Congruency in this study does not presumably arise out of audibility factors, since the Congruent and Incongruent sounds were equated for rms, but listeners' expectations for sounds in a familiar scene. Thus, it likely reflects the influence of cognitive factors such as selective attention and ability to inhibit irrelevant stimuli. Overall, the results suggest that elderly listeners may face a dual challenge in dealing with complex auditory environments due to the loss of high frequency hearing and, potentially, a general decline in cognitive abilities.

This points to the possible effects of rehabilitative strategies such as hearing aids. Restoring high frequency cues to persons with hearing loss, hearing aids are likely to improve overall performance (resulting in lower thresholds for both Congruent and Incongruent sounds). However, because IA also seems dependent on cognitive abilities, at this point, it is not clear whether it may be affected by the use of assistive hearing devices. Recent studies of speech perception have demonstrated interactions between audibility and cognition (Lunner, Rudner, & Rönnberg, 2009; Lunner & Sundewall-Thorén, 2007), suggesting that amplification may influence IA in hearing impaired individuals. This hypothesis, however, remains to be tested.

Further research can demonstrate specific cognitive factors involved in the IA effect and provide a basis for developing rehabilitation strategies to improve environmental sound perception in natural settings for the elderly. The results presented here point to ways to dissociate the auditory and cognitive processes. In this instance, the diminished audiometric capabilities of the elderly affected the threshold of the psychometric function, whereas the effect of cognitive decline may have resulted in a decreased slope of the psychometric function for Congruent sounds. This indicates methods that can assist in experimentally isolating these factors for future study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Nabeel Rahman for assistance with listener testing.

This work was supported by a Merit Review Training Grant from the Department of Veterans Affairs Research Service, VA File # 06-12-00446 to Brian Gygi. Partial support was also provided by the National Institute of Deafness and Other Communication Disorders (DC0008676) to Valeriy Shafiro.

References

Andringa, T.; Grootel, Mv. Predicting listeners' reports of environmental sounds. Paper presented at the 19th International Congress on Acoustics; Madrid, Spain. 2007.

Aydelott J, Leech R, Crinion J. Normal Adult Aging and the Contextual Influences Affecting Speech and Meaningful Sound Perception. Trends in Amplification. 2010; 14(4):218–232. [PubMed: 21307006]

Ballas JA, Howard JH. Interpreting the language of environmental sounds. Environment & Behavior. 1987; 19(1):91–114.

Ballas JA, Mullins T. Effects of context on the identification of everyday sounds. Human Performance. 1991; 4(3):199–219.

Cherry C. Some experiments on the recognition of speech with one and with two ears. Journal of the Acoustical Society of America. 1953; 26:975–979.

Cohen, J. Statistical Power Analysis for the Behavioral Sciences. 2. Taylor & Francis; 1988.

Divenyi PL, Haupt KM. Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation. Ear and Hearing. 1997; 18(3):189–201. [PubMed: 9201454]

Dobreva MS, O'Neill WE, Paige GD. Influence of aging on human sound localization. Journal of Neurophysiology. 105(5):2471–2486. [PubMed: 21368004]

Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of outpatients for the clinician. Journal of Psychiatric Research. 1975; 12:189–198. [PubMed: 1202204]

Friederici AD, Steinhauer K, Frisch S. Lexical integration: sequential effects of syntactic and semantic information. Memory and Cognition. 1999; 27(3):438–453. [PubMed: 10355234]

Ganis G, Kutas M. An electrophysiological study of scene effects on object identification. Cognitive Brain Research. 2003; 16:123–144. [PubMed: 12668221]

Gilbert JC. Memory loss in senescence. Journal of Abnormal and Social Psychology. 1941; 36:73–86.

Gold JI, Law CT, Connolly P, Bennur S. Relationships between the threshold and slope of psychometric and neurometric functions during perceptual learning: implications for neuronal pooling. Journal of Neurophysiology. 2010; 103(1):140–154. [PubMed: 19864439]

Greene MR, Oliva A. Recognition of natural scenes from global properties: Seeing the forest without representing the trees. Cognitive Psychology. 2009; 58(2):137–76. [PubMed: 18762289]

Gygi B, Kidd GR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. Journal of the Acoustical Society of America. 2004; 115(3):1252–1265. [PubMed: 15058346]

Gygi B, Kidd GR, Watson CS. Similarity and Categorization of Environmental Sounds. Perception and Psychophysics. 2007; 69(6):839–855. [PubMed: 18018965]

Gygi, B.; Shafiro, V. From Signal to Substance and Back: Insights from Environmental Sound Research to Auditory Display Design. In: Ystad, S.; Aramaki, M.; Kronland-Martinet, R.; Jensen, K., editors. Auditory Display: 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18–22, 2009. Revised Papers. Vol. 5954. Berlin / Heidelberg: Springer; 2010. p. 306-329.

Gygi B, Shafiro V. The incongruency advantage for environmental sounds presented in natural auditory scenes. Journal of Experimental Psychology: Human Perception and Performance, Vol. 2011; 37(2):551–565.

Hasher, L.; Zacks, RT. Working memory, comprehension, and aging: A review and a new view. In: Bower, GH., editor. Advances in research and theory: The psychology of learning and motivation. Vol. 22. San Diego, CA: Academic Press; 1988. p. 193-225.

Hasher L, Zacks RT, Rahhal TA. Timing, instructions, and inhibitory control: some missing factors in the age and memory debate. Gerontology. 1999; 45(6):355–357. [PubMed: 10559658]

Helmholtz, Hv. On the Sensations of Tone. Ellis, AJ., translator. London: Longmans, Green, and Co; 1875. Vol. English text based upon the 3rd edition, 1870

Humes LE. Aging and speech communication. The ASHA Leader. 2008; 13(5):10–13. 33.

Humes LE, Lee JH, Coughlin MP. Auditory measures of selective and divided attention in young and older adults using single-talker competition. The Journal of the Acoustical Society of America. 2006; 120(5):2926. [PubMed: 17139749]

Humes LE, Watson BU, Christensen LA, Cokely CG, Halling DC, Lee L. Factors associated with individual differences in clinical measures of speech recognition among the elderly. Journal of Speech & Hearing Research. 1994; 37(2):465–474. [PubMed: 8028328]

International Standards Organization (ISO). Acoustics-Statistical Distribution of Hearing Thresholds as a Function of Age, ISO-7029. Basel, Switzerland: ISO; 2000.

Kramer AF, Humphrey DG, Larish JF, Logan GD, Strayer DL. Aging and inhibition: beyond a unitary view of inhibitory processing in attention. Psychology of Aging. 1994; 9(4):491–512.

Lass NJ, Eastman SK, Parrish WC, KAS, Ralph D. Listeners' identification of environmental sounds. Perceptual & Motor Skills. 1982; 55(1):75–78. [PubMed: 7133923]

Lawrence DM, Banks WP. Accuracy of recognition memory for common sounds. Bulletin of the Psychonomic Society. 1973; 1(5A):298–300.

Leech R, Gygi B, Aydelott J, Dick F. Informational factors in identifying environmental sounds in natural auditory scenes. The Journal of the Acoustical Society of America. 2009; 126(6):3147. [PubMed: 20000928]

Leek MR, Watson CS. Learning to detect auditory pattern components. The Journal of the Acoustical Society of America. 1984; 76(4):1037–1044. [PubMed: 6501698]

Li L, Daneman M, Qi JG, Schneider BA. Does the Information Content of an Irrelevant Source Differentially Affect Spoken Word Recognition in Younger and Older Adults. Journal of Experimental Psychology: Human Perception & Performance. 2004; 30(6):1077–1091. [PubMed: 15584816]

Lunner T, Rudner M, Rönnberg J. Cognition and hearing aids. Scandinavian Journal of Psychology. 2009; 50(5):395–403. [PubMed: 19778387]

Lunner T, Sundewall-Thorén E. Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid. Journal of the American Academy of Audiology. 2007; 18(7):604–617. [PubMed: 18236647]

Meihong W, Huahui L, Zhiling H, Xinchi X, Jingyu L, Xihong W, et al. Effects of aging on the ability to benefit from prior knowledge of message content in masked speech recognition. Speech Communication. 2012; 54(4):529–542.

Moore, BCJ. Cochlear hearing loss. London: Whurr Publishers; 1998.

Niessen ME, Van Maanen L, Andringa TC. Disambiguating sound through context. International Journal of Semantic Computing. 2008; 2(3):327–341.

Palmer SE. The effects of contextual scenes on the identification of objects. Memory & Cognition. 1975; 3:519–526. [PubMed: 24203874]

Pichora-Fuller MK. Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. International Journal of Audiology. 2003; 42:S59–67. [PubMed: 12918611]

Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen and remember speech in noise. Journal of the Acoustical Society of America. 1995; 97(1):593–608. [PubMed: 7860836]

Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychological Review. 1996; 103(3):403–428. [PubMed: 8759042]

Schneider BA, Daneman M, Murphy DR. Speech comprehension difficulties in older adults: cognitive slowing or age-related changes in hearing? Psychology of Aging. 2005; 20(2):261–271.

Schneider BA, Daneman M, Murphy DR, See SK. Listening to discourse in distracting settings: the effects of aging. Psychology of Aging. 2000; 15(1):110–125.

Schneider, BA.; Pichora-Fuller, K.; Daneman, M.; Gordon-Salant, S.; Frisina, RD.; Popper, AN., et al. The Aging Auditory System. Vol. 34. Springer; New York: 2010. Effects of Senescent Changes in Audition and Cognition on Spoken Language Comprehension; p. 167-210.

Shafiro V, Gygi B. How to select stimuli for environmental sound research and where to find them. Behavior Research Methods, Instruments & Computers. 2004; 36(4 part 2):590–598.

Shafiro V, Gygi B, Cheng MY, Mulvey M, Holmes B. Perception of speech and environmental sounds in cochlear implant patients. The Journal of the Acoustical Society of America. 2008; 123(5):3303.

Sheft S, Shafiro V, Lorenzi C, McMullen R, Farrell C. Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception. Ear and Hearing. 2012; 33(6):709–720. [PubMed: 22790319]

Singh G, Pichora-Fuller MK, Schneider BA. Age-related differences in auditory spatial attention depend on task switching complexity. The Journal of the Acoustical Society of America. 2010; 127(3):1810–1810.

Truax, B. Acoustic Communication. Westport, Connecticut: Ablex Publishing; 2001.

Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling and goodness-of-fit. Perception and Psychophysics. 2001; 63(8):1293–1313. [PubMed: 11800458]

Wichmann FA, Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. Perception & Psychophysics. 2001; 63(8):1314–1329. [PubMed: 11800459]

Appendix

The supplemental material contains some examples of stimuli used in the experiments described here. Racing.wav is a recording of an auto race scene. Auto racein2Cd9.wav is the auto race scene mixed with a congruent sound. Auto raceout2A9.wav is the auto race scene mixed with an incongruent sound. Both of these are at −9 So/Sc. See if you can identify the target sounds. Listen to Congruent sound.wav and Incongruent sound.wav to hear the target sounds in the clear.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

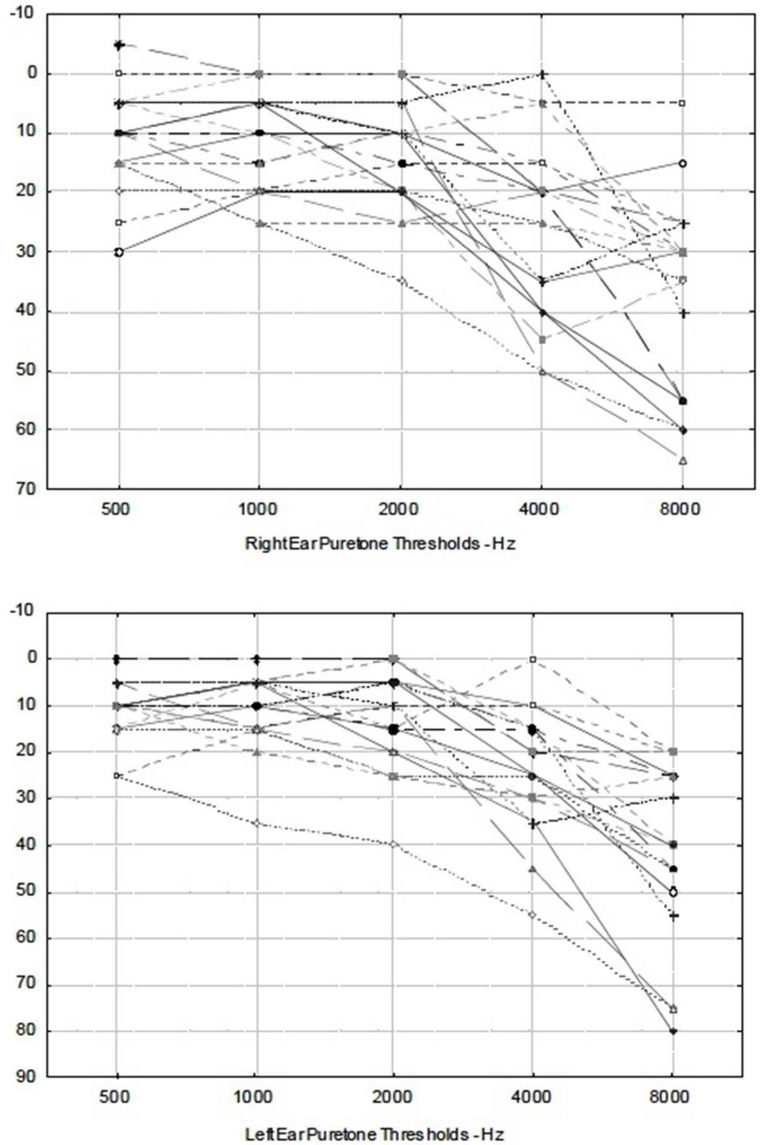

Figure 1.
Right and Left-Ear Puretone Thresholds for the Participants in Experiments 1 & 2.

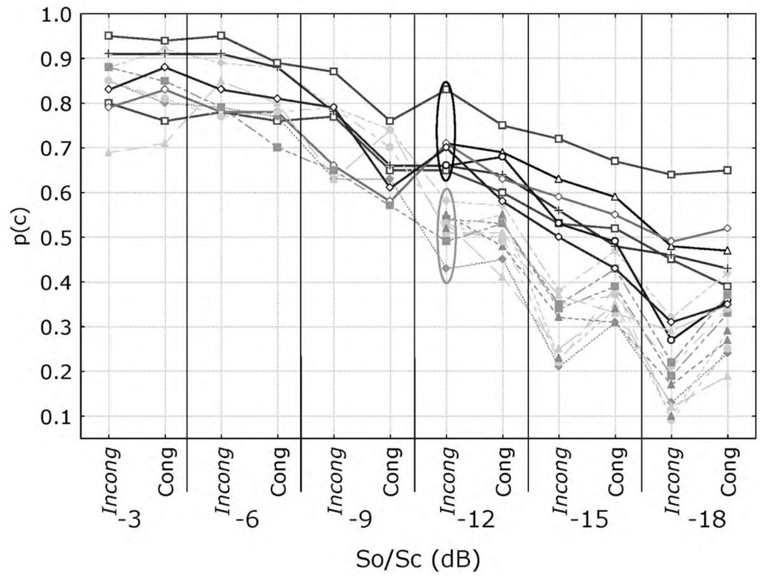

Figure 2.

Individual listener p(c) from Experiment 1a&1b by So/Sc and Congruency condition. The data for So/Sc −12 to −18 dB are from Experiment 1a, and the data for So/Sc −3 to −9 dB are from Experiment 1b. The ovals contain the higher- and lower-performing subjects for the analysis in Results for Experiment 1a. In addition, the symbols for lower-performing subjects' data points are all filled, both gray and black, and the lines are dotted or dashed, whereas the higher-performing subjects' datapoints are not filled and the lines are solid. The performance of the groups is compared in Figure 4.

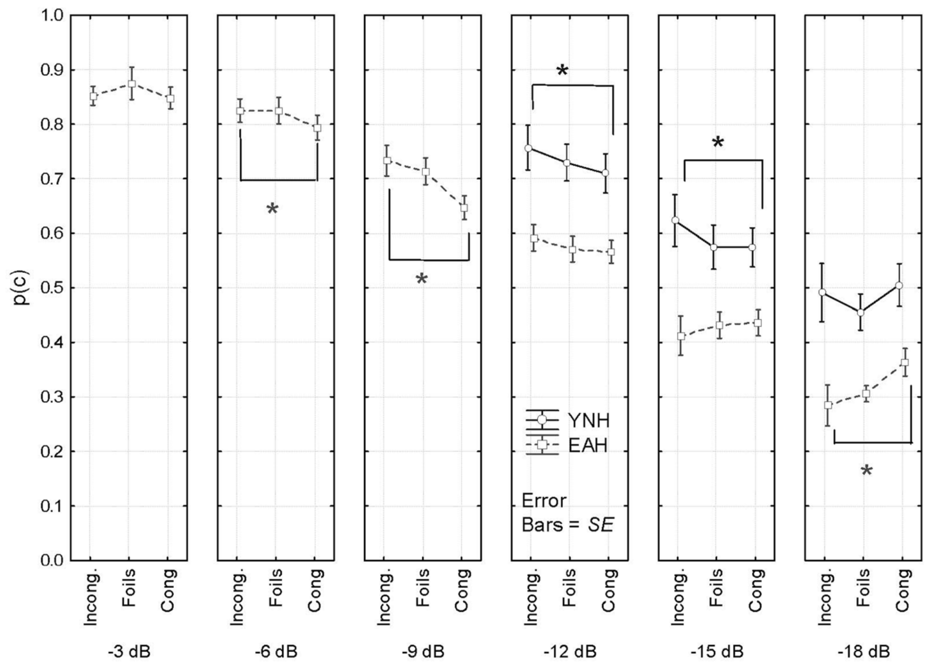

Figure 3.

Mean performance of elderly by So/Sc and Congruency, plotted along with the data for experienced YNH from Gygi & Shafiro (2011) in the same conditions. Significant differences at the $p<0.05$ level between the Incongruent and Congruent conditions are marked with an asterisk; a double asterisk denotes $p<0.01$. Note that at the −18 dB So/Sc the difference for the EAH is an advantage for the Congruent sound/scene pairs. Overall, the data patterns for the Experienced YNH resemble that for the EAH, except at ~3 dB lower So/Sc. Error bars represent ± 1 Standard Error.

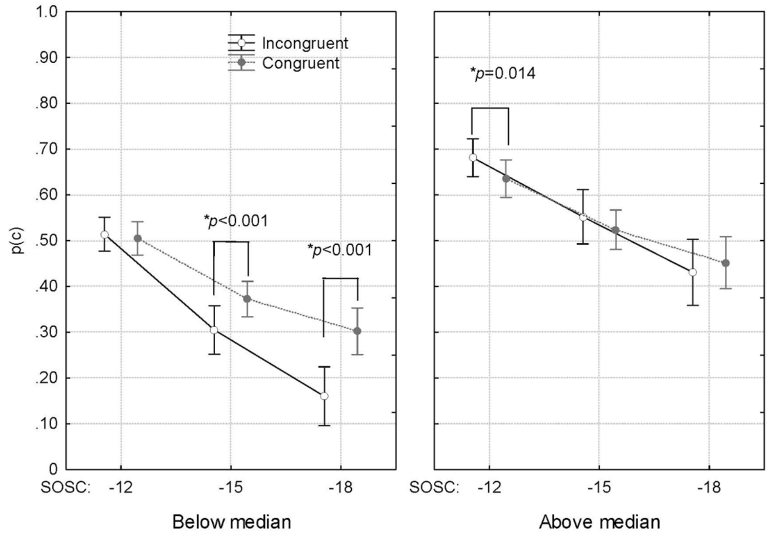

Figure 4.

Performance of the better-performing participants (right panel) versus the lower-performing (left-panel) group from Experiment 1a, categorized based on their scores in the −12 dB So/Sc condition. The bars indicate ± 1 Standard Error. The below-median listeners exhibited a significant Congruency advantage (based on Fisher LSD post hoc test) at −18 and −15 So/Sc, whereas the above-median listeners showed an Incongruency Advantage at −12 dB. Further discussion in text.

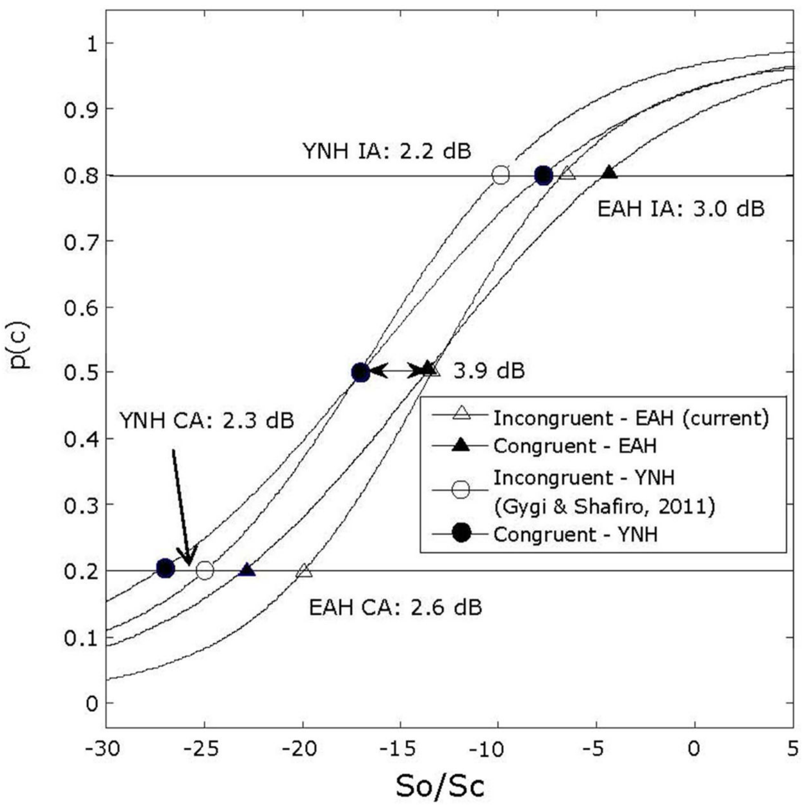

Figure 5.

Psychometric functions for the ENH listeners compared to the YNH from Gygi & Shafiro (2011). The lines represent cuts through the psychometric functions at 0.2 and 0.8 p(c). Where the threshold in the Incongruent condition is lower than the threshold in the Congruent, that is an Incongruency Advantage (IA). Where the threshold for the Congruent condition is the lower one, it is a Congruency Advantage (CA). Both groups of listeners showed an IA at 0.8 p(c) and a CA at 0.2 p(c). The 95% confidence intervals for these functions are included in Table 1.

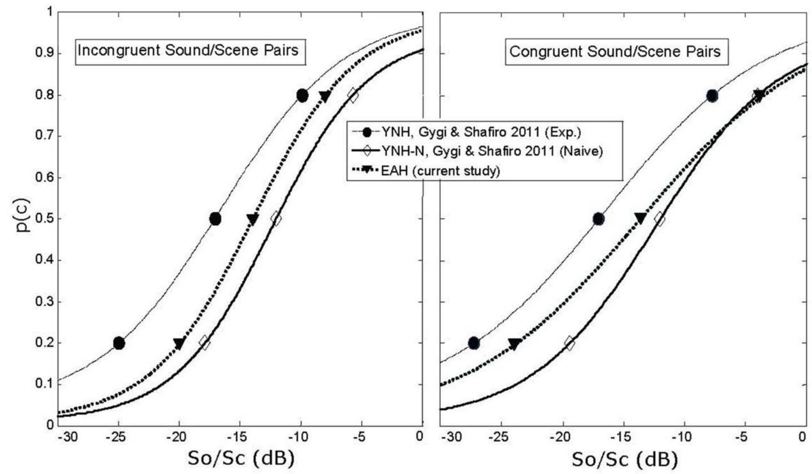

Figure 6.

Psychometric functions for the ENH plotted alongside those for the YNH Experienced and Naïve listeners from Experiments 2 and 3 from Gygi & Shafiro (2011), grouped by Congruency rather than together as in Figure 5, to highlight the group differences. Note that the vertical scale is the same for both figures and so was omitted from the right figure.

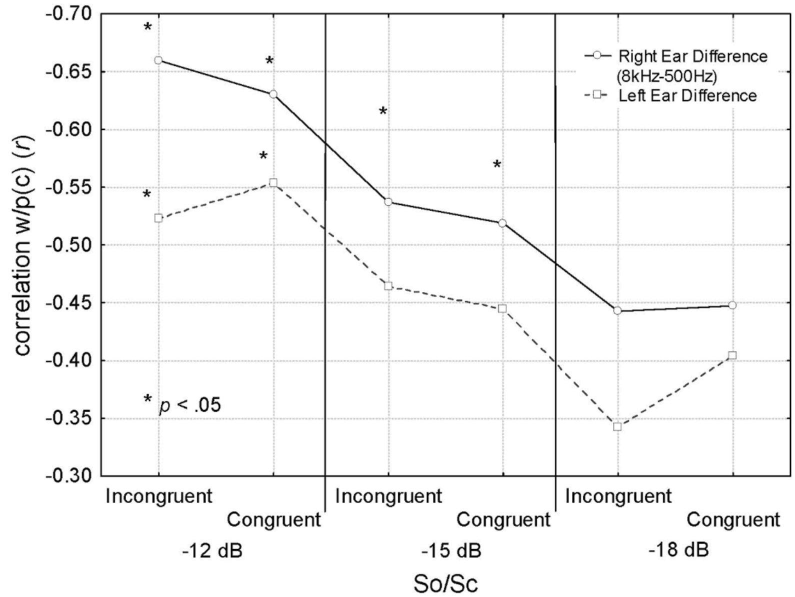

Figure 7.
Correlations of the difference between the 8kHz and 500 Hz audiogram threshold in each Congruency condition with p(c) in the low So/Sc levels from Experiment 1. Significant correlations are marked with an asterisk (*).

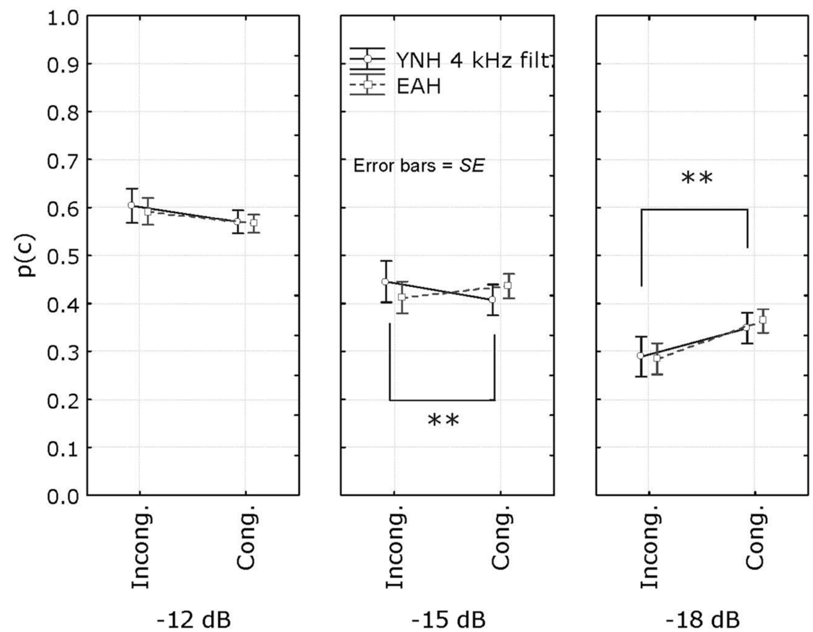

Figure 8.

Results for the YNH listeners in Experiment 2 for stimuli filtered at 4kHz, paired with the EAH results at the same So/Sc. A double asterisk (**) notes differences for YNH-4kHz across Congruence conditions at a particular So/Sc which are significant at $p < .01$. The error bars represent ± 1 Standard Error.

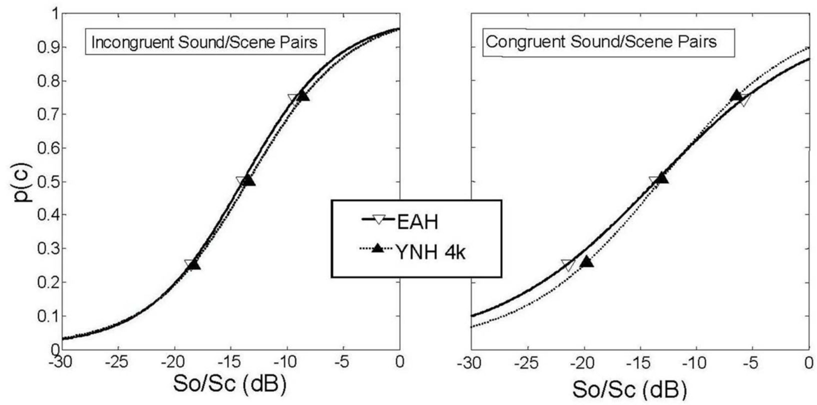

Figure 9.
The psychometric functions for the EAH listeners from Experiment 1a and the YNH listeners who heard stimuli filtered at 4 kHz, grouped by Congruency.

Table 1

Thresholds at the 0.8, 0.5 and 0.2 p(c) isoperformance levels, the 95% Confidence Interval (CI) for those thresholds, and the slopes in both Incongruent and Congruent conditions for the subject groups discussed in this paper. EAH are the experienced elderly listeners, and the YNH-4k are the experienced young listeners presented with stimuli that were low pass filtered at 4 kHz. For comparison purposes, thresholds and slopes from Gygi and Shafiro (2011) are included: YNH and YNH-N are the young normal experienced and naïve listeners, respectively. For the thresholds, a lower number means a lower threshold. For the slopes, a lower number means a shallow slope.

		EAH	**YNH-4k**	**YNH**	**YNH-N**
Threshold @ 0.80 p(c)/.95 CI	Incongruent	−7.55±0.58	−8.13±0.88	−10.25	−5.54
	Congruent	−4.51±0.84	−4.70±1.25	−7.98	−4.65
Threshold @ 0.50 p(c)/.95 CI	Incongruent	−13.78±0.34	−13.93±0.33	−17.71	−11.86
	Congruent	−13.58±0.40	−13.11±0.69	−17.59	−12
Threshold @ 0.20 p(c)/.95 CI	Incongruent	−20.02±0.45	−19.73±0.53	−24.96	−17.74
	Congruent	−22.64±0.60	−21.53±0.92	−27.19	−18.86
Slope	Incongruent	.0562	.0597	.0478	.0624
	Congruent	.0352	.0412	.0361	.0536