Psychology and Aging

Frequent False Hearing by Older Adults: The Role of Age Differences in Metacognition

Chad S. Rogers, Larry L. Jacoby, and Mitchell S. Sommers

Online First Publication, December 12, 2011. doi: 10.1037/a0026231

Psychology and Aging

0882-7974/11/$12.00 DOI: 10.1037/a0026231

Frequent False Hearing by Older Adults: The Role of Age Differences in Metacognition

Chad S. Rogers, Larry L. Jacoby, and Mitchell S. Sommers
Washington University in St. Louis

In two experiments testing age differences in the subjective experience of listening, which we call meta-audition, young and older adults were first trained to learn pairs of semantic associates. Following training, both groups were tested on identification of words presented in noise, with the critical manipulation being whether the target item was congruent, incongruent, or neutral with respect to prior training. Results of both experiments revealed that older adults compared to young adults were more prone to "false hearing," defined as mistaken high confidence in the accuracy of perception when a spoken word had been misperceived. These results were obtained even when performance was equated across age groups on control items by reducing the noise level for older adults. Such false hearing is shown to reflect older adults' heavier reliance on context. Findings suggest that older adults' greater ability to benefit from semantic context reflects their bias to respond consistently with the context, rather than their greater skill in using context. Procedures employed are unique in measuring the subjective experience of hearing as well as its accuracy. Both theoretical and applied implications of the findings are discussed. Convergence of results with those showing higher false memory, and false seeing are interpreted as showing that older adults are less able to constrain their processing in ways that are optimal for performance of a current task. That lessened constraint may be associated with decline in frontal-lobe functioning.

Keywords: aging, context effects, speech perception, metacognition, meta-audition

During a hike in coastal Michigan, I (C.R.) approached a large fir tree to take a picture. A small songbird that was nesting in the tree swooped toward me aggressively, defending its territory. When later relating this incident to my family, I told them that I was attacked by a lark. My grandmother was alarmed, asking "How are you possibly all right?!" After some discussion, it became clear that she was absolutely certain that she had "heard" me say that I had been attacked by a *shark*. My grandmother's error is an example of what we refer to as "false hearing"—a high-confidence, subjective experience of having actually "heard" a misperceived word (e.g., shark). This article presents evidence that false hearing is more common among older adults than young adults. We argue that measures of subjective experience, as reflected by false hearing, are a critical yet underutilized assessment tool in audition and that age differences in subjective experience provide novel insight into the mechanisms that mediate perceptual experience.

When perceiving a spoken word in naturalistic listening situations, listeners can base their perceptual experience on two distinct sources of information: sensation and context (e.g., Nittrouer & Boothroyd, 1990). Sensation refers to the acoustic and phonetic characteristics of the word as processed by the peripheral auditory system. Context refers to the mental and environmental circumstances within which the word is perceived. In the above example, sensory information refers to phonetic cues, including formant frequencies, voice-onset times, burst frequencies, and other phonetic information that listeners use to identify the linguistic content of speech signals. Context, on the other hand, refers to the information contained in the sentence prior to presentation of the target word. Given that sharks are known to attack people, in the above example, sensory and contextual information are incongruent in that the sensory information strongly suggests "lark" whereas the contextual information strongly suggests "shark." Based on evidence soon described, we predicted that older adults would be more likely to falsely hear words presented in a misleading context than young adults.

While subjective experience measures such as confidence ratings have not been strongly emphasized in audition research, they have frequently been used in investigations of metacognition. Of particular interest is the extent to which confidence ratings distinguish between correct and incorrect responses. We employ a measure called *monitoring resolution* to assess the correlation between confidence in identification of a spoken word and identification accuracy. As suggested by monitoring and control frameworks of metacognition (e.g., Koriat & Goldsmith, 1996; Nelson & Narens, 1990), such resolution is important because people control their actions on the basis of their confidence. The goal of this work is to integrate findings from metacognition, memory, and audition with the aim of understanding age differences in false

Chad S. Rogers, Larry L. Jacoby, and Mitchell S. Sommers, Department of Psychology, Washington University in St. Louis.

Chad S. Rogers is now with the Volen National Center for Complex Systems at Brandeis University.

This research was supported by National Institute on Aging Grant AG000030 to Larry L. Jacoby. We thank Anna Dinndorf, Carole Jacoby, Emily Norwood, Tim Bono, and Tiffany Adams for their assistance with data collection.

Correspondence concerning this article should be addressed to Chad S. Rogers, Volen National Center for Complex Systems, Brandeis University, Waltham, MA, 02454. E-mail: rogers@brandeis.edu

hearing. We use the term "meta-audition" to refer to the metacognitive aspect of audition. Just as studies of metamemory have helped further understanding of memory processes (e.g., Dunlosky & Metcalfe, 2009; Koriat & Goldsmith, 1996), research on meta-audition has the potential to advance understanding of auditory processing. We highlight further similarities between age differences in meta-audition and age differences in metamemory.

We begin by considering research on differential use of contextual information by young and older adults. Prior work aimed at understanding age differences in the use of context (e.g., Nittrouer & Boothroyd, 1990) has been limited to situations in which context supports correct identification. Importantly, those experiments have not examined situations in which context could give rise to false hearing, such as the "shark/lark" example described above. Next, we review parallels between the processes underlying age differences in metamemory and meta-audition by relating false hearing to false memory. Finally, two experiments are described that assessed context effects on meta-audition in young and older adults.

Aging and The Facilitative Use of Context in Speech Perception

As people grow older, the relative contributions to speech perception made by context typically increases (e.g., Nittrouer & Boothroyd, 1990). In particular, older adults rely more on top-down information such as semantic context to compensate for hearing impairment (Wingfield, Tun, & McCoy, 2005). When young and older adults are compared under degraded listening conditions (e.g., with moderate to high levels of background noise present), age differences in spoken word identification diminish significantly in the presence of supportive semantic context (Dubno, Ahlstrom, & Horwitz, 2000; Hutchinson, 1989; Pichora-Fuller, Schneider, & Daneman, 1995; Sommers & Danielson, 1999). Given these findings, some have suggested that older adults are more skilled than young adults at using context, resulting from heavier reliance on context when listening in degraded conditions in daily life (Pichora-Fuller, 2008). Sommers and Danielson (1999) also presented evidence showing that the addition of context reduced age differences in lexical discrimination (Sommers, 1996). They argued that context constrains activation to a smaller set of candidate words, and that because of older adults' deficit in ability to inhibit alternative responses (e.g., Hasher & Zacks, 1988), context is more beneficial for older than young adults.

An alternative account is that reliance on context produces a bias effect rather than serving to increase discrimination. A bias effect would show itself by increasing false hearing when context and the sensory signal were incongruent as well as by increasing correct hearing when the sensory signal and context were congruent. In contrast, models that attribute age differences in effects of context to differences in enhanced discrimination alone would predict correct hearing without increasing false hearing (e.g., NAM, Luce & Pisoni, 1998, and PARSYN, Luce et al., 2000).

Reliance on contextual information is generally adaptive because context may only rarely be misleading. Consequently, older adults' greater reliance on context is generally useful as a means of compensating for hearing deficits. We characterize reliance on context as focusing on a larger unit of the *word-in-context* rather than focusing on the *individual word.* These two levels serve as qualitatively different bases for auditory judgments, analogous to the letter and word levels in investigations of visual perception aimed at the word superiority effect (e.g., Reicher, 1969; Wheeler, 1970). For hearing, focusing at the word level is a more effortful method for parsing heard messages than is focusing at the word-in-context level, but is necessary for correctly identifying words that are spoken in an incongruent context. Such focus was called a "close look" by Bruner (1957), who noted that participants must constrain their perception to specific features of an object in order to avoid top-down biases that result in perceptual illusions.

Supportive context has been shown to reduce perceptual effort in older listeners (McCoy et al., 2005), thus, older adults might be less able than young adults to effortfully focus attention at the word level so as to avoid false hearing. This may be true even when they are warned that context will often be misleading, and when ability to identify words without supportive context has been equated by reducing background noise for older adults. Similar hypotheses have been supported by findings in the visual domain reported by Jacoby, Rogers, Bishara, and Shimizu (2011). In their procedure, older and young adults had to identify briefly flashed words. They found that if a word was preceded by a misleading prime, older adults were more likely to report subjectively "seeing" the primed word (i.e., false seeing).

For the same reasons that older adults show greater false memory and false seeing in tasks where the fluency of a response can be misleading (e.g., Hay & Jacoby, 1999; Jacoby, Bishara, Hessels, & Toth, 2005a; Jacoby, Rogers, Bishara, & Shimizu, 2011), older adults were expected to show greater false hearing from reliance on misleading context. Jacoby and Rhodes (2006) review experiments demonstrating that older adults are much more prone to false memory than are young adults, and describe those findings in terms of a dual-process model of memory that distinguishes between recollection and accessibility bias. Recollection is described as a consciously controlled, effortful basis for responding that is tightly constrained by retrieval cues. In contrast, accessibility bias is a less effortful, more automatic basis for responding that reflects more global factors such as prior experience in the form of habits and context. They argue that the controlled processes necessary for supporting recollection diminish with age, rendering bias effects more influential with the result that the probability of false memory is increased. Just as avoiding false memory requires a close look at the past (constraining processing in ways required for recollection) avoiding false seeing and false hearing require a close look or close listen to the present.

Present Experiments

The present experiments investigated age differences in false hearing with procedures akin to those used by Jacoby and colleagues (e.g., Hay & Jacoby, 1999; Jacoby et al., 2005a) to investigate false memory. Congruent and incongruent associative contexts were created by utilizing a cue-target training procedure that required participants to learn semantically related pairs of words (e.g., BARN-HAY). At test, participants listened to the cue word (e.g., BARN) presented in the clear and then listened to a word masked by white noise. For congruent trials, the word in noise was the same as the trained target (HAY). For incongruent trials, the word in noise was a phonological neighbor that formed a minimal pair (see Luce & Pisoni, 1998) with the trained target

(PAY). For baseline trials (the control condition), the word in noise was unrelated to the training target (FUN). Experiment 1 manipulated the level of noise during presentation of the target word and utilized a two-alternative, forced-choice test (2AFC, HAY/PAY). After selecting an alternative, participants indicated how confident they were that they had identified the word correctly.

In Experiment 2, the false hearing procedure was generalized to a more naturalistic listening situation. Instead of white noise, Experiment 2 utilized a 6-talker babble noise-masking procedure and an open-set, cued identification task, in which the participant said aloud the word that was presented in the noise. To control for age-related differences in speech perception ability, the noise level for each participant was set to their 50% speech reception threshold (SRT; ASHA, 1988). This was important to allow certainty that any age differences in false hearing that were observed did not result from age-related sensory differences in hearing.

In both experiments, participants' meta-audition was assessed by analyzing mean confidence data, high confidence errors, and monitoring resolution. Confidence ratings have been commonly employed in studies of aging and metamemory (e.g., Jacoby, Wahlheim, Rhodes, Daniels, & Rogers, 2010; Kelley & Sahakyan, 2003; Lovelace & Marsh, 1985; Perfect & Stollery, 1993), and were chosen as to assess how well participants subjectively thought they were hearing. As mentioned earlier, resolution measures the extent to which a person's confidence discriminates between correct and incorrect responses. Resolution was assessed using Goodman-Kruskal (Goodman & Kruskal, 1954) gamma correlations at the item-level to examine the correspondence between confidence and accuracy (see Nelson, 1984, for a discussion of the advantages of using gamma as a measure of metamemory). Like a Pearson's correlation coefficient, a gamma correlation ranges from -1 to $+1$, where the absolute value reflects the degree of association, and the direction of the association is indicated by positive or negative values. A strong gamma correlation indicates that confidence strongly distinguishes between correct and incorrect responses, whereas a weak gamma correlation implies little association between confidence and accuracy.

For resolution, we expected to find an interaction between age and context type. For congruent contexts, we expected the resolution of confidence judgments to be *higher* for older than for young adults. When context is congruent, reliance on context provides a valid basis for accuracy, as does reliance on the audibility of the word. In contrast, for incongruent items we expected the resolution of confidence judgments to be *lower* for older than for young adults. In the incongruent condition context provides invalid cues for accuracy; thus, older adults' greater reliance on context should result in poorer resolution for their confidence judgments. An interaction of this sort would provide strong evidence for a qualitative difference in the basis for confidence used by older versus young adults.

Experiment 1

Method

Participants. Sixteen undergraduate students were recruited through the Washington University subject pool and received either $15 or course credit for their participation. These young participants ranged in age from 18 to 22 years ($M = 19.75$, $SD = 1.18$). Sixteen older adults were recruited through the Washington University Older Adult subject pool. These older participants ranged in age from 65–82 years ($M = 75.63$, $SD = 4.49$), and received $15 for their participation. The mean score on the Vocabulary subtest of the Shipley Institute of Living Scale (Shipley, 1967) was lower for young participants ($M = 33.75$, $SD = 2.77$) than for older participants ($M = 35.81$, $SD = 2.64$), $t(30) = 2.15$, $p < .05$. All participants reported normal or corrected-to-normal vision.

Pure Tone Audiometric thresholds were obtained for all participants, and these thresholds were used to screen for hearing loss. Participants were tested using an audiometer in a double-walled sound-attenuating booth. None of the older adults or young adults had thresholds exceeding 25dB HL for frequencies of 500, 1,000, and 2,000 Hz.

Materials and design. Signal-to-noise ratio (SNR; -10 or -15) and trial type (Congruent, Incongruent, or Baseline) were manipulated within participants. A total of 72 three-word sets including one cue word (e.g., BARN), one associatively related monosyllabic target word (e.g., HAY), and one nonassociatively related monosyllabic alternate word that was phonologically confusable with the target word (e.g., PAY) were generated using the Washington University Neighborhood Database (Sommers, 2000) to create the congruent and incongruent trials. The three-word sets were divided into four groups of 18, which were balanced for word frequency and phonological confusability. These groups were rotated across participants through each of the combinations of congruent/incongruent trial types and SNR levels (e.g., congruent -10, congruent -15, incongruent -10, and incongruent -15). A total of 36 three-word sets were used for constructing baseline trials. Those three-word sets contained a cue word (e.g., CLOUD) and two monosyllabic words that were not associatively related to the cue, but were phonologically confusable with one another (e.g., FUN, RUN). Two groups of 18 of these baseline word-sets were balanced for word frequencies and phonological confusability, so that they could be rotated across participants through the two SNR levels.

The auditory stimuli were spoken versions of the above word sets recorded at 11,025 Hz using a 16-bit Digital-to-Analog converter with a Shure microphone in a double-walled sound attenuating booth. Words were spoken by a female speaker with a standard American dialect. Root-mean-square (RMS) amplitude of the stimuli was equated. Stimuli masked by noise were generated by taking the clear speech file (65dB SPL) and mixing it with a corresponding white noise file (75dB SPL for the -10 SNR condition and 80 dB SPL for the -15 SNR condition) using Adobe Audition v1.5 (Adobe, 2004).

Procedure

Training phase. The procedure for the experiment was broken into two phases: the training phase and the perception test phase. The purpose of the training phase was to create a strong context by training the cue-target pairs to a high level of accuracy. During the training phase, participants were seated 18 in. away from the front of a computer screen and learned a series of word pairs that they were told to remember for a later memory test. For each pair, the cue word (e.g., BARN) was presented on screen, and

then 100 ms later was presented aurally via headphones. Fifty ms later, the associatively related target word (e.g., HAY) was presented, adjacent to the cue word, aurally and visually in the same fashion. Both words presented visually remained on the screen for the entirety of aural presentation. Each cue-target pair was presented a total of five times. Pairs were presented in random order, with the limitation that all 72 pairs were presented once before any pair was presented an additional time.

The final component of the training phase was a 72-item cued-recall test to assess training. On each trial, the cue word was presented visually and aurally, but with a question mark following the word (e.g., BARN−?). Participants had five seconds to provide the target word and were encouraged to guess if they did not know. After a response was provided, or five seconds elapsed, the target word was presented visually adjacent to the cue word (e.g., BARN–HAY), and the word was played over the headphones. All participants correctly recalled 80% or more of these words.

Perception test phase. During the 108-trial perception test phase, there were three different trial types: congruent, incongruent, and baseline trials. There were 36 of each trial type, half of which had target words presented at an SNR of −10, the other half at −15. Order of conditions in the perceptual test phase was randomized for each participant, with the limitation that no more than three trials of a given type were presented consecutively. Participants were informed that they would again be hearing a series of cue-target pairs, but that during this portion of the experiment the target word would be masked by noise. Participants were told that after the word in noise was played, two words would appear on the screen. Their task was to pick which of the two words was presented in the noise by saying the word aloud (i.e., 2AFC). Participants were warned that some of the pairs in the perception test phase would be the same as the pairs in the training phase (e.g., BARN–HAY) but that some of the pairs would be different (e.g., BARN–PAY), and because of this they should only respond on the basis of what they heard in the noise, not on what they had learned earlier. This last point was printed in capital letters on the computer screen and was emphasized by the experimenter when recapitulating the instructions.

After providing an identification judgment, participants were instructed to indicate how confident they were that they had provided the correct response. As in the 2AFC perception test, the participants gave their rating aloud and the experimenter recorded the response. The 50-point scale for this judgment ranged from 50–100. Participants were encouraged to use the full range of the scale. The scale began at 50 because with the 2AFC test, participants had a 50% chance of providing the correct response based on pure guessing. As with the identification judgments, participants were instructed to make their confidence judgments only on the basis of what they heard in the noise.

After participants received all instructions for the perceptual test phase, they were asked to explain the procedure in their own words. Participants' reports had to include (a) the identification judgment, (b) the confidence rating, and (c) the misleading nature of context. The instructor verbally repeated instructions and questioned participants until each participant's procedure report was complete. All participants' procedure reports were complete before the beginning of the perceptual test phase.

The timing for each trial was as follows: 200 ms before the first member of a pair (the cue) was presented over the headphones, a single asterisk "*" was presented visually in the top center portion of the screen until the offset of the aurally presented word. Following a 1,000 ms interstimulus interval, two asterisks "**" were presented visually in the top center of the computer screen; 200 ms later the target word, masked by noise, was presented aurally. The asterisks were used so that participants would have a visual indication of which word was being played over the headphones, but were offset so that they did not distract the participants while the word was being played.

Results and Discussion

Unless otherwise specified, only effects that were found to be significant at $\alpha < .05$ significance level and that were not involved in a higher-order interaction are reported. When Mauchly's test of sphericity was significant, the Greenhouse-Geyser correction for *MSE* and degrees of freedom was used.[1] Furthermore, when Levene's test for equality of variances was significant during post hoc *t* tests, the degrees of freedom were corrected.

Hit rates. Identification accuracy was measured as the proportion of trials on which participants correctly identified the word in noise (hits). Figure 1 shows that while young adults had more hits than older adults on baseline trials, older adults had more hits than young adults on congruent trials. This finding is consistent with the notion that older adults effectively utilize context to compensate for age-related hearing loss (Hutchinson, 1989; Pichora-Fuller et al., 1995; Pichora-Fuller, 2008; Sommers & Danielson, 1999; Wingfield et al., 2005). However, in our novel incongruent condition, it is clear that this contextual facilitation came at a cost: older adults had fewer hits than young adults on incongruent trials. This pattern of greater contextual facilitation for older adults on congruent trials and greater contextual interference on incongruent trials was consistent across the −10 and −15 SNR conditions, indicating a stronger reliance upon context for older adults than young adults.

To confirm the statistical reliability of these findings, hit rates were analyzed using a 2 (age: young, older) × 2 (SNR: −10, −15) × 3 (trial type: congruent, baseline, incongruent) mixed-model analysis of variance (ANOVA). The age x trial type interaction was significant, $F(1.85, 55.59) = 7.71$, $MSE = .37$, $p < .001$, $\eta_p^2 = .20$. Post hoc *F* tests applying the Bonferroni Type I error correction revealed that across SNRs older adults showed more hits on congruent trials ($M = .83$, $SD = .14$) than did young adults ($M = .72$, $SD = .16$), $F(1, 30) = 4.04$, $p < .05$, but fewer hits on incongruent trials ($M = .24$, $SD = .17$) than did the young adults ($M = .43$, $SD = .16$), $F(1, 30) = 10.97$, $p < .01$. Baseline performance was lower for older adults ($M = .55$, $SD = .084$) than for young adults ($M = .62$, $SD = .085$), $F(1, 30) = 5.02$, $p < .05$, indicating an age group difference in context-free speech percep-

[1] The Greenhouse-Geyser correction is a widely used and accepted adjustment for the increased rate of Type I errors that can occur when data are not spherical. This adjustment is robust when used with balanced designs (Keselman & Rogan, 1980; Keselman, 1998). A more conservative multivariate analysis of variance (MANOVA) approach, suggested by O'Brien and Kaisser (1985), was also conducted because MANOVA does not assume sphericity. In these studies, all tests that were significant using the Greenhouse-Geyser correction for ANOVA were also significant using MANOVA, with our manipulations included as fixed effects.

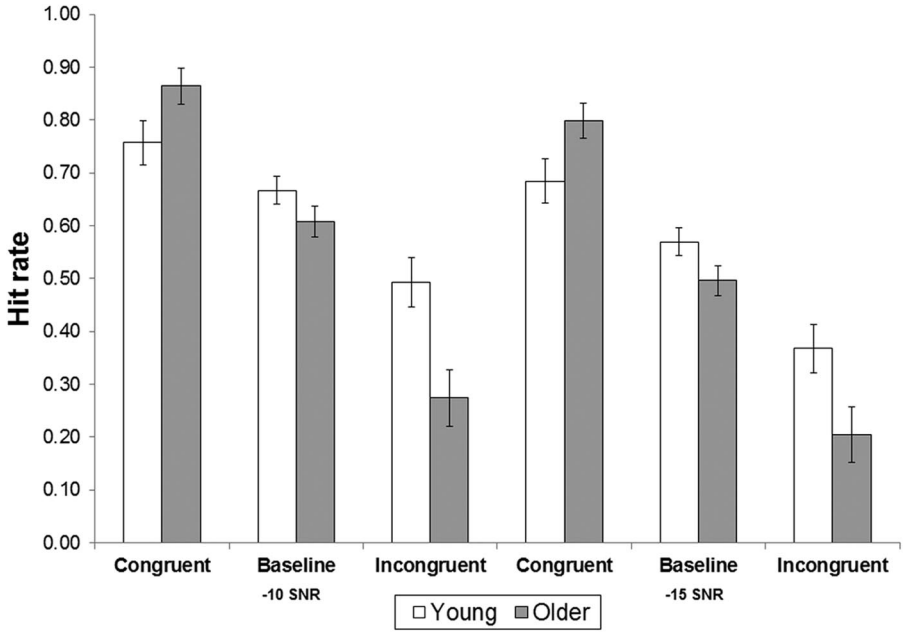

Figure 1. Hit rates for identifications made in Experiment 1. SNR denotes signal-to-noise ratio. Error bars represent standard errors.

tion ability. As expected, participants had greater hits at better SNRs, as revealed by a significant main effect of SNR, $F(1, 60) = 31.28$, $MSE = .386$, $p < .001$, $\eta_p^2 = .51$.

Confidence data. For the incongruent condition, choosing the alternative favored by context (i.e., the incorrect response) and holding high confidence in its selection served as a measure of false hearing. If participants were aware of cases in which they had failed to identify the word presented in noise and instead responded on the basis of context, then a choice predicated upon context might be considered a low-confidence "best guess." In contrast, if reliance on context resulted in the chosen word being subjectively experienced as "heard," then confidence in context-favored responses should be high. We expected older adults, compared to young adults, to show greater confidence in words favored by context in both the congruent and incongruent test conditions.

The confidence pattern depicted in Figure 2 shows the mean confidence rating ascribed to responses that were favored by context (i.e., congruent hits and incongruent false alarms). The baseline condition serves as a reference point for correct identification made without prior context. The most striking finding that emerges from an examination of Figure 2 is a "V-shaped" function indicating that older adults were very confident when choosing a response favored by context. In contrast, young adults' confidence judgments were minimally influenced by context. Young adults were more confident in their choices made in the baseline condition than were older adults. The 2 (age: young, older) $\times$ 2 (SNR: -10, -15) $\times$ 3 (trial type: congruent, baseline, incongruent) mixed-model ANOVA on mean confidence ratings revealed a significant 3-way interaction of trial type, SNR, and age, $F(1.74, 60) = 3.40$, $MSE = 68.97$, $p < .05$, $\eta_p^2 = .10$. Separate analyses done for the two SNRs revealed that in the -10 condition, the age x trial type interaction was highly significant, $F(1.58, 60) = 25.803$, $MSE = 851.11$, $p < .001$, $\eta_p^2 = .46$, but was smaller in the -15 condition, $F(1.21, 60) = 6.76$, $MSE = 396.54$, $p < .01$, $\eta_p^2 = .18$. We attribute this attenuation of the age x trial type interaction to potential floor effects on baseline trials for the -15 SNR condition.

High confidence errors (Dramatic false hearing). Erroneously selecting the alternative favored by context in the incongruent condition and expressing l00% confidence that the selected word was the one presented in noise we define as "dramatic false hearing." Older adults ($M = .20$, $SD = .27$) tended to be more

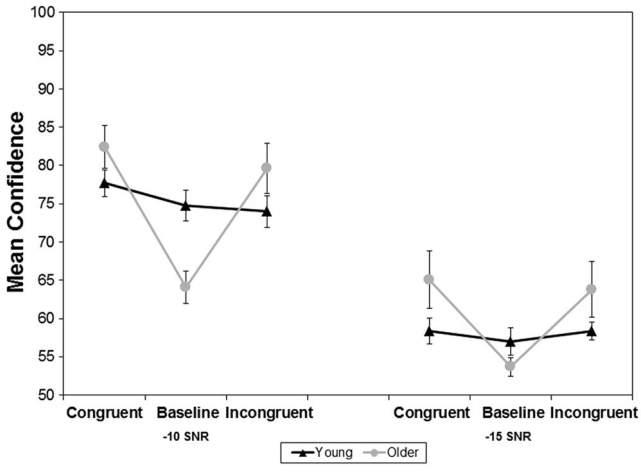

Figure 2. Confidence in responses favored by context in Experiment 1. Confidence in hits is plotted for congruent and baseline trials. Confidence in false alarms is plotted for incongruent trials. SNR denotes signal-to-noise ratio. Error bars represent standard errors.

likely than young adults ($M = .05$, $SD = .05$) to exhibit dramatic false hearing across SNRs, as shown by a main effect of age that approached significance, $F(1, 30) = 3.60$, $MSE = 72.25$, $p < .07$, $\eta_p^2 = .11$. Dramatic false hearing was also less likely to occur in the −15 ($M = .04$, $SD = .14$) than in the −10 ($M = .12$, $SD = .25$) SNR condition, as indicated by a significant main effect of SNR, $F(1, 30) = 9.23$, $MSE = 33.06$, $p < .01$, $\eta_p^2 = .24$. In Experiment 2, age differences in dramatic false hearing were further examined with a more powerful design and a greater number of participants.

Resolution. Recall that resolution is a measure of metacognitive monitoring that assesses the extent to which confidence in a response can discriminate whether the response was correct or not. As stated in the introduction, we expected the resolution to be higher for older than for young adults on congruent trials, and to be lower for older adults than for young adults on incongruent trials. This interaction would show that age groups differ in their bases for responding, with young adults more likely to respond on the basis of sensory information and older adults being more reliant upon context.

Resolution was assessed using gamma correlations. When participants used only one point on a confidence scale or achieve either 0% or 100% accuracy, a gamma correlation could not be calculated. In the case of two young adults and six older adults, participants' gamma correlations could not be calculated for the above reasons, and were excluded from analysis. Figure 3 shows the resolution data from the remaining 14 young adults and 10 older adults.[2]

The resolution data presented in Figure 3 reveal the predicted interaction, where older adults had better monitoring than young adults on congruent trials, but poorer monitoring than young adults on incongruent trials. The 2 (age: young, older) × 2 (SNR: −10, −15) × 3 (trial type: congruent, baseline, incongruent) mixed-model ANOVA on the gamma correlations revealed that the predicted age x trial type interaction was highly significant, $F(2, 44) = 10.47$, $MSE = 2.44$, $p < .001$, $\eta_p^2 = .32$. Post hoc F tests using the Bonferroni correction revealed age group differences to be significant on congruent trials, $F(1, 22) = 20.82$, $MSE = 1.05$, $p < .001$, $\eta_p^2 = .49$, and incongruent trials, $F(1, 22) = 10.55$, $MSE = 1.06$, $p < .01$, $\eta_p^2 = .32$, but not baseline trials, $F(1, 22) = 2.81$, $p > .10$, *ns*. There was also a main effect of SNR, $F(1, 22) = 4.44$, $MSE = 1.18$, $p < .05$, $\eta_p^2 = .19$, which suggests that resolution was poorer in the −15 SNR condition.

The resolution results provide strong evidence of a qualitative difference between young and older adults in bases for responding. In the congruent condition, both sensory and context information were valid, and older adults showed a high positive correspondence between confidence and accuracy. In the incongruent condition, context was invalid, and older adults showed a strong *negative* correspondence between confidence and accuracy. This negative resolution implies that the more confident the listener was in his or her response, the more likely he or she was to be incorrect. The magnitude of these correlations is important as an indication of the extent to which older adults relied on context. Although young adults did show negative gammas in the incongruent condition, they were not as strong as those of the older adults.

Summary. Experiment 1 employed a 2AFC procedure and demonstrated that older adults were more reliant on context than were young adults when selecting an item as having been presented in noise. Older adults also were more prone to show dramatic false hearing, by being maximally confident in their erroneous selection of an item in the incongruent context condition. The finding of a significant age x trial type interaction for resolution provided strong evidence that young and older adults relied on qualitatively different bases for judging confidence.

The larger increase in correct responding in combination with the larger increase in false hearing shown by older adults provides evidence that the greater advantage of providing facilitative context for older adults found in prior experiments (Dubno et al., 2000; Hutchinson, 1989; Pichora-Fuller et al., 1995; Sommers & Danielson, 1999) results from a bias effect that is akin to that responsible for higher false seeing (Jacoby, et al., 2011) and false remembering by older adults (e.g., Hay & Jacoby, 1999).

Experiment 2

Experiment 2 included several changes from Experiment 1. The most important change is that a noise-adjustment procedure we call *titration* was used to equate performance of young and older adults on control trials. Because young adults had a higher baseline hit rate than older adults in both −10 and −15 SNR listening conditions, some might argue that the greater false hearing found for older adults in Experiment 1 might simply reflect age-related deficits in hearing. This difference in baseline accuracy does not truly compromise conclusions from Experiment 1 because of the general lack of interactions between SNR and age. However, in Experiment 2, the SNR was adjusted for each participant to a level that would produce approximately 50% correct identification, often referred to as the speech reception threshold (SRT, ASHA, 1988). SRT is a common clinical measure used to assess an individual's ability to understand speech in noise. Our goal was to show that results from Experiment 1 could be replicated even when young and older adults were equated in their performance on baseline items (cf. for false memory, Jacoby et al., 2005a).

Another change is Experiment 2 gave participants the opportunity to act on the basis of their subjective experience. Participants were given a volunteer/withhold response option (Kelley & Sahakyan, 2003; Koriat & Goldsmith, 1996) after producing

[2] Gammas were most often incalculable in the −15 SNR condition, where some participants reported the lowest level of confidence on every baseline trial (i.e., 50%), and/or chose the response favored by context on every congruent or incongruent trial (i.e., 100% accuracy on congruent trials, 0% accuracy on incongruent trials). To confirm the reliability of the results and minimize the number of participants excluded, the −10 SNR condition was separately analyzed. The same pattern as shown in Figure 3 was obtained: with only three older participants excluded, a 2 (age: young, older) × 3 (trial type: congruent, baseline, incongruent) mixed-model repeated measures ANOVA revealed a significant age x trial type interaction, $F(1.38, 3.30) = 9.18$, $MSE = 2.29$, $p < .001$, $\eta_p^2 = .25$. Post hoc F tests applying the Bonferroni correction revealed that older adults showed better resolution than young adults on congruent trials (Young $M = .39$, Older $M = .81$), $F(1, 27) = 10.97$, $MSE = 1.23$, $p < .01$, $\eta_p^2 = .29$, poorer resolution than young adults on incongruent trials (Young $M = .05$, Older $M = -.66$), $F(1, 27) = 8.18$, $MSE = 3.63$, $p < .01$, $\eta_p^2 = .23$, and trended toward poorer resolution than young adults on baseline trials (Young $M = .44$, Older $M = .22$), $F(1, 27) = 2.73$, $p > .10$, *ns*.

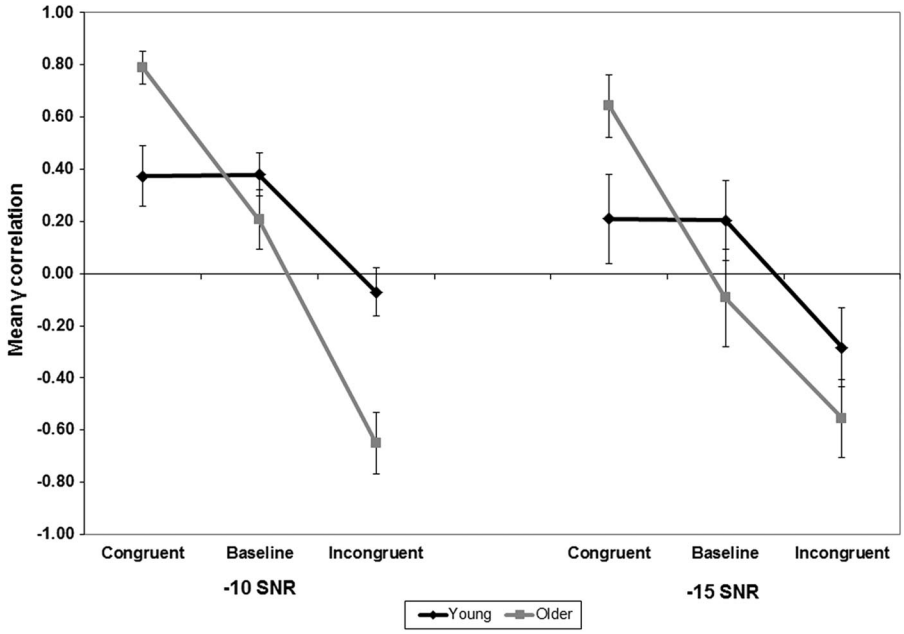

Figure 3. Gamma (γ) correlation data from Experiment 1. Values above the zero line correspond to a positive relationship between confidence and accuracy (good monitoring), whereas values below the zero line correspond to a negative relationship between confidence and accuracy (poor monitoring). SNR denotes signal-to-noise ratio. Error bars represent standard errors.

an identification and confidence response. Participants were told that the computer was keeping score by giving one point for correct responses that were volunteered, and penalizing one point for incorrect responses that were volunteered. Participants were told to attempt to maximize their score. Further, they were instructed that they could improve their score by withholding responses that they produced that were of uncertain accuracy. Based on the resolution data from Experiment 1, we expected older adults to be less able to successfully monitor their accuracy in the incongruent condition. Such monitoring is critical for determining which responses to withhold (Koriat & Goldsmith, 1996), and, therefore, we expected older adults to benefit less from the option of withholding responses than would young adults. In particular, compared with young adults, we expected older adults to show an increased probability of false hearing as well as increased volunteering words that were falsely heard. In naturalistic settings, older adults do have the ability to withhold responses and could enhance their hearing performance by doing so. Consequently, lessened ability to withhold erroneously heard items is of applied as well as of theoretical interest.

A third change was a shift from a closed-set (e.g., 2AFC) to an open-set identification procedure to simulate more naturalistic hearing situations. Unlike 2AFC, in typical listening situations individuals are not presented with a list of potential responses. In Experiment 2, instead of two viable alternatives appearing on the screen, a single question mark "?" appeared. Participants were instructed to say the word aloud that they believed was presented in the noise. As in Experiment 1, participants followed their identification attempt with a confidence judgment, this time ranging from 0–100. The type of noise was changed from white noise to 6-talker babble to more closely simulate real-world listening in noise situations.

Method

Participants. Participants were recruited from the same pools as in Experiment 1. Twenty-five young adult participants ranged in age from 18−22 years ($M = 19.28$, $SD = 0.94$). Twenty-five older adults ranged in age from 65–87 years ($M = 75.40$, $SD = 6.16$). Shipley vocabulary scores were lower for young ($M = 33.16$, $SD = 2.66$) than for older participants ($M = 35.36$, $SD = 2.29$), $t(48) = 3.14$, $p < .005$.

Materials and design. In Experiment 2, trial type (congruent, incongruent, and baseline) was the only within-subjects manipulation. There was not a within-participants manipulation of SNR, as in Experiment 1. Instead, the SNR was set to each individual's SRT.

Experiment 2 used 6-talker babble instead of white noise to mask speech. The babble was captured from the Iowa Audiovisual Speech Perception Laserdisc (Tyler, Preece, & Tye-Murray, 1986) using a 16-bit converter at a sampling rate of 44,100 Hz. Because there was no SNR manipulation, set sizes for congruent, incongruent, and baseline items increased from 18 to 36. The list of 72 congruent/incongruent word sets from Experiment 1 was separated into two groups that were balanced for frequency and phonological confusability, and were then rotated across participants. The two groups of 18 baseline word sets from Experiment 1 were combined into a single 36-item list. This resulted in two forms of the experiment, in which each group of congruent/

incongruent word sets occurred equally often in both congruent and incongruent conditions.

Procedure

Setting the SNR. Experiment 2 used a modification of the American Speech-Language-Hearing Association's recommended procedure (ASHA, 1988) for obtaining a SRT, and then used the SNR for each participant's SRT as the SNR for the perceptual test phase. The ASHA procedure is used to measure an individual's threshold for speech, as opposed to their ability to hear a pure tone. The mean SRT obtained for older adults was higher ($M = 6.92$, $SD = 3.81$) than for young adults ($M = 2.96$, $SD = 1.94$), $t(48) = 4.16$, $p < .001$.

Training phase. With one exception, the training phase of Experiment 2 was identical to the one used in Experiment 1. The exception was that four additional pairs were added to the training list to be used as congruent/incongruent practice items during the perceptual test phase.

Perceptual test phase. The 108-trial perceptual test phase of Experiment 2 was similar to the perceptual test phase from Experiment 1. However, instead of the 2AFC recognition task used in Experiment 1, an open-set cued recall task was used. For congruent and baseline trials, correct responses were coded as hits and incorrect responses were coded as misses. For incongruent trials, correct responses were coded as hits, the invalid responses favored by context were coded as false alarms, and all other responses were coded as misses. Following their response, participants were prompted to make a judgment that reflected how confident they were that they had correctly identified the word in the noise, with a scale ranging from 0 to 100, and then were given the option to volunteer/withhold their response from scoring. Before participants began the perceptual test, they were given six practice trials consisting of two instances of each trial type.

Results and Discussion

Hit rate. The identification (hit rate) data from Experiment 2, plotted in Figure 4, show that older adults had more hits on congruent trials than did young adults, but age groups did not differ on baseline or incongruent trials. The 2 (age: young, older) × 3 (trial type: congruent, baseline, incongruent) repeated-measures mixed model ANOVA on hit rates revealed a significant age x trial type interaction, $F(1.47, 70.59) = 5.88$, $MSE = .084$, $p < .01$, $\eta_p^2 = .11$. Planned age comparisons, based on the results of Experiment 1, revealed that older adults had greater hits on congruent trials relative to young adults, $t(39.90) = 4.35$, $p < .001$, but also indicated that older and young adults did not significantly differ on baseline or incongruent trials, $ts < 1$, *ns*. Most importantly, the lack of an age group difference on baseline trials served as a successful check for the titration procedure used in Experiment 2.

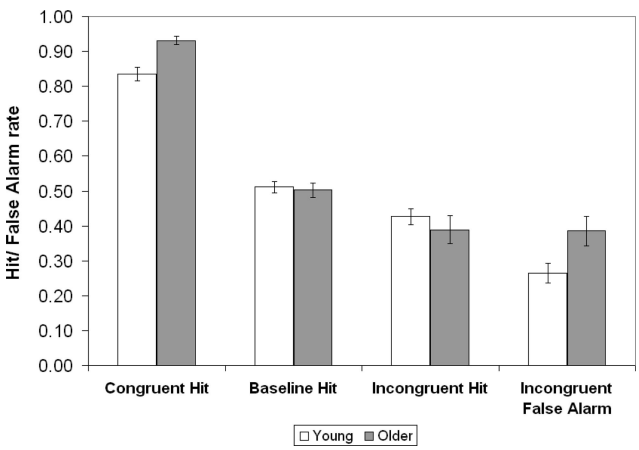

Figure 4. Hit and false alarm rates for identifications made in Experiment 2. Error bars represent standard errors.

The lack of age group differences for incongruent hits may come as a surprise given the results of Experiment 1. However, the primary interest in false hearing is the incongruent false alarm rate (also plotted in Figure 4). Older adults ($M = .39$, $SD = .21$) were significantly more likely than young adults ($M = .26$, $SD = .14$) to produce an incongruent false alarm, $t(41.80) = 2.44$, $p < .02$.

Confidence data. The mean confidence ratings assigned to congruent hits, baseline hits, and incongruent false alarms are plotted in Figure 5. That figure shows that unlike Experiment 1, young adults also exhibited the "V" pattern that implies use of context as a basis for their meta-audition. However, the pattern of higher confidence for older adults than young adults on trials in which the response was favored by context was replicated, as revealed by a significant interaction of trial type and age, $F(1.38, 66.23) = 14.26$, $MSE = 2047.87$, $p < .001$, $\eta_p^2 = .23$. Planned age comparisons, based on the results of Experiment 1, revealed that older adults were more confident in their congruent hits than were young adults, $t(48) = 4.99$, $p < .001$, as well as in their incongruent false alarms, $t(48) = 3.54$, $p < .001$. Age groups did not differ significantly in their confidence for baseline hits, $t < 1$, *ns*. Age groups also did not significantly differ in their congruent, baseline, or incongruent misses, or incongruent hits, all $ts < 1.20$, *ns*.

Dramatic false hearing. Even after the SNR was set for each individual as a method for controlling for age group differences in word intelligibility, older adults ($M = .27$, $SD = .18$) were almost four times more likely to exhibit dramatic false hearing (i.e., incongruent false alarms with 100% confidence) than young adults ($M = .07$, $SD = .13$), $t(48) = 4.46$, $p < .001$. Assuming that confidence ratings of 100% indicate the subjective experience of hearing a word, on more than one quarter of incongruent trials, older adults falsely "heard" a word that was not presented.

Resolution. Monitoring resolution was again measured using gamma correlations in Experiment 2. Figure 6 shows that consistent with the results of Experiment 1, older, compared to younger, adults showed better monitoring on congruent trials, equivalent monitoring on baseline trials and poorer monitoring on incongruent trials. The statistical reliability of older adults' greater reliance on contextual information in monitoring was confirmed by a significant age x trial type interaction, $F(1.40, 58.69) = 4.43$, $MSE = .57$, $p < .01$, $\eta_p^2 = .10$. Because of perfect accuracy on congruent trials, five older adults and one young adult were excluded from the ANOVA. Planned age comparisons revealed that older and young adults did not differ in their resolution on baseline trials, $t(48) < 1$, *ns*, older adults showed better resolution on congruent trials, $t(42) = 2.18$, $p < .05$, and poorer resolution on incongruent trials $t(48) = 2.3$, $p < .05$.

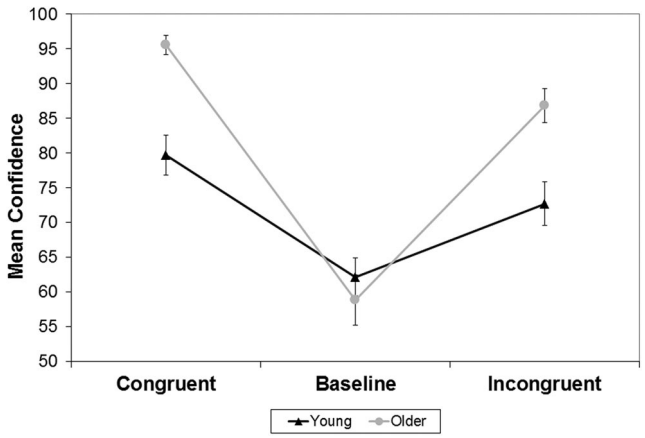

Figure 5. Confidence in responses favored by context in Experiment 2. Confidence in hits is plotted for congruent and baseline trials. Confidence in false alarms favored by context is plotted for incongruent trials. Error bars represent standard errors.

Metacognitive control. A main goal of Experiment 2 was to investigate how older and young adults use metacognitive control to *act* on the basis of their subjective experience of hearing. These types of decisions more closely simulate naturalistic settings, where listeners are given the option to not respond. To assess participants' metacognitive control three different measures were analyzed: (a) the rate of responses volunteered, (b) the rate of volunteered incongruent false alarms, and (c) the difference in proportion correct when considering only responses that were volunteered compared to proportion correct overall.

Koriat and Goldsmith (1996) showed that the rate of responses volunteered is strongly predicted by confidence in those responses. Thus, planned age comparisons were used, based on the confidence data. These comparisons showed that older adults ($M = .94$, $SD = .09$) were significantly more likely than young adults ($M = .77$, $SD = .17$) to volunteer on congruent trials, $t(36.06) = 4.18$, $p < .001$. Likewise on incongruent trials, older adults ($M = .75$, $SD = .21$) were significantly more likely to volunteer than young adults ($M = .59$, $SD = .16$), $t(48) = 3.01$, $p < .01$. On baseline trials, older adults ($M = .04$, $SD = .16$) also tended to be more likely to volunteer than were young adults ($M = .53$ vs. $SD = .32$), but the age group difference was not significant, $t(34.90) = 1.50$, $p = .14$. Of greater concern to false hearing was the rate of volunteered incongruent false alarms. The joint occurrence of committing an incongruent false alarm and volunteering it to be scored was greater in older ($M = .35$, $SD = .20$) than young adults ($M = .20$, $SD = .14$), $t(42.30) = 3.21$, $p < .01$.

Participants' ability to improve their accuracy using the volunteer/withhold response option was examined. To assess this we, like others in the metacognition literature (Jacoby et al., 2005a; Kelley & Sahakyan, 2003; Koriat & Goldsmith, 1996), compared the overall hit rates [i.e., p(hit)] to the hit rates for only the subset of trials that were volunteered [i.e., p(hit|volunteered)]. Overall hit rate was then subtracted from the conditional hit rate [i.e., p(hit|volunteered)−p(hit)], to assess gains in accuracy resulting from use of the volunteer/withhold response option. Because older adults' overall hit rates were near ceiling on congruent trials ($M = .93$), our discussion is limited to baseline and incongruent trials. A 2 (age: young, older) × 2 (trial type: baseline, incongruent) mixed-model ANOVA assessing difference scores in overall and conditional hit rates revealed a significant age x trial type interaction, $F(1, 48) = 4.23$, $MSE = .075$, $p < .05$, $\eta_p^2 = .08$. Post hoc F tests revealed that on baseline trials, young ($M = .16$, $SD = .09$) and older adults ($M = .17$, $SD = .15$) demonstrated equivalent increments in accuracy using the volunteer/withhold response option, $F < 1$, *ns*. However, on incongruent trials, young adults were able to increase their accuracy ($M = .08$, $SD = .15$), but older adults were not ($M = -.03$, $SD = .10$), $F(1, 48) = 12.42$, $p < .001$. Indeed, older adults' conditionalized hit rate trended toward being lower than their overall hit rate [p(hit|volunteered) $M = .39$ versus p(hit) $M = .36$], although the trend was not significant by a 1-sample t test of the difference score against a test value of 0, $t(24) = 1.32$, $p < .20$. This result, along with the age-related increase in volunteered incongruent false alarms provides evidence that monitoring deficits for older adults on incongruent trials led to poorer control.

Summary. The results of Experiment 2 replicated the results of Experiment 1 by showing that older adults were more reliant on context as a basis for responding than were young adults. This greater reliance on context was revealed by effects on identification and confidence. Age differences remained even after overall hearing differences between young and older adults were controlled using the SRT procedure. The addition of a volunteer/withhold response showed that older adults acted on the basis of their subjective experience. Just as found for false memory (Jacoby et al., 2005a; Kelley & Sahakyan, 2003), older adults did not increase the accuracy of their hearing when given the option to withhold responses. In terms of metacognitive monitoring and control frameworks (Koriat & Goldsmith, 1996; Nelson & Narens, 1990), older adults' reliance upon context led to an age deficit in both monitoring and control on incongruent trials.

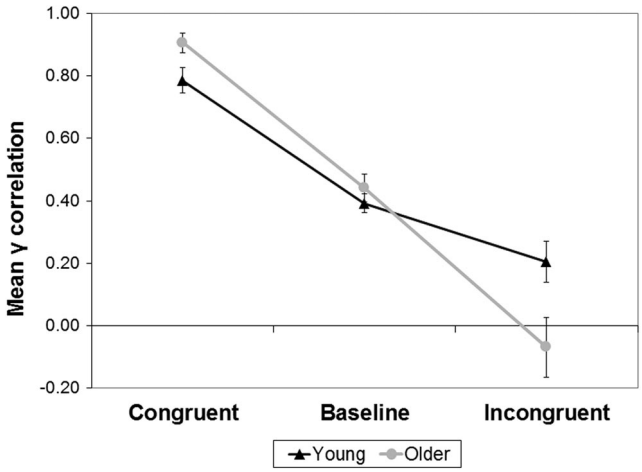

Figure 6. Gamma (γ) correlation data from Experiment 2. Values above the zero line correspond to a positive relationship between confidence and accuracy (good monitoring), whereas values below the zero line correspond to a negative relationship between confidence and accuracy (poor monitoring). Error bars represent standard errors.

General Discussion

To summarize our findings, Experiments 1 and 2 found that older adults were more likely than young adults to use context as a basis for responding. Doing so led to greater accuracy for older adults when context was facilitative, but produced greater false hearing when context was incongruent. Use of an open-set procedure where SNR was titrated to control for age-related changes in hearing ability reduced the age difference in correct responding in the incongruent condition to the extent that the difference was no longer significant; however, large age differences in false hearing persisted. When given the option to volunteer or withhold their responses on incongruent trials, older adults were not able to increase their accuracy, whereas young adults did so. It is notable that for baseline items, young and older adults did not differ in their ability to improve their accuracy by withholding responses.

The above results converge on the conclusion that context and sensory information are two qualitatively different bases for hearing, and that older adults are more reliant upon context for accuracy and confidence than are young adults.

As described earlier, resolution refers to the correspondence between confidence and accuracy at the item level (Nelson, 1984). High resolution is adaptive for hearing and would be reflected by greater relative doubt for responses that were less likely to be correct. Results from both experiments revealed that older adults' resolution of confidence judgments was *higher* than that of young adults when context was congruent but *lower* than young adults' resolution when context was incongruent. Importantly, young and older adults did not differ in resolution in the baseline condition. These results provide strong evidence for qualitative differences in bases for confidence. The lack of a difference in resolution in the baseline condition indicates that age differences in other conditions were not the result of a difference in use of the confidence scale; nor were they due to a general difference in monitoring ability.

For both age groups, providing context enhanced performance on congruent trials compared to the baseline condition but produced poorer performance on incongruent trials. This increase in both hits and false alarms provides evidence that the effect of providing context was, at least, partially due to an influence on bias. Earlier findings showing that facilitative context diminishes or eliminates age differences in hearing (e.g., Hutchinson, 1989; for a recent review, see Pichora-Fuller, 2008) can be explained as a result of older adults' heavier reliance upon context. The prevailing account for this age-related increase in the benefit of semantic context has focused on how context serves to constrain potential alternative responses. For example, Sommers and Danielson (1999) suggested that context functions to reduce the initial set of lexical candidates by activating only those items that are *both* semantically consistent with the context and phonologically consistent with the auditory signal. Enhanced performance for older adults produced by facilitative context can be accounted for by models of this sort. However, these constraint-based accounts (e.g., Sommers & Danielson, 1999) could only accommodate the accuracy data from the incongruent condition by assuming that bias toward context can prevent the correct alternative from entering into the set of potential response candidates.

Convergence Between False Hearing, False Seeing, and False Memory

Just as older adults are more likely to falsely hear than are young adults, they are also more likely to falsely see (Jacoby et al., 2011) and falsely remember (e.g., Hay & Jacoby, 1999). We propose that these findings reflect a general deficit in cognitive control, as opposed to specific deficits in vision, hearing, and memory. In vision, Jacoby et al. (2011) found that when an uppercase identity prime preceded a briefly flashed lowercase target (e.g., DART–dart), older adults had better visual identification than young adults. When the uppercase prime was misleading (e.g., DIRT–dart), older adults were more likely than young adults to falsely see the primed word. In memory, Jacoby et al. (2005a) found dramatic false remembering by older adults in an incongruent context condition using procedures analogous to those used in this study in conjunction with a priming manipulation. In that study, older adults seldom took advantage of the opportunity to withhold responses ("pass") as a means of avoiding false recall, whereas young adults did take advantage of that option. These age differences were found even though older and young adults were equated on baseline performance. Kelley and Sahakyan (2003) also found that older adults were less likely to increase their accuracy by withholding responses.

The above age differences in false remembering were interpreted in terms of a dual-process model of memory that is analogous to the model that we propose to explain age differences in audition and vision. By that dual-process model, conscious recollection and more automatic influences of memory serve as alternative bases for responding. Recollection is a more effortful basis for responding that relies on cue specification to tightly constrain retrieval to the particular prior event in which a target item occurred, whereas automatic influences reflect a more global basis for responding (e.g., Jacoby et al., 2005a). Jacoby, Shimizu, Velanova, and Rhodes (2005b) provide evidence older adults did not adjust cue-specification processes to the details of a recognition memory task in ways done by young adults, but instead relied on a more global basis of responding. We explain age differences in hearing and seeing in terms of a similar distinction by suggesting that young adults are more capable of constraining their response to perceived sensory information (e.g., carefully listening and closely looking), whereas older adults respond using a more global basis that includes contextual information. In the false memory, false hearing, and false seeing paradigms, high confidence in an erroneous response may reflect a misattribution stemming from reliance on a fluency heuristic, which treats the ease with which a response comes to mind as grounds for confidence in its accuracy (e.g., Jacoby & Dallas, 1981). In other words, fluency creates the illusion of a veridical perceptual experience.

A dual-process framework that postulates the use of a fluency heuristic as a basis for confidence can account for the relationship between confidence and accuracy in both congruent and incongruent context conditions in our study. Such a framework acknowledges that there are invalid as well as valid bases for confidence. The importance of validity of the cues used for confidence judgments has been acknowledged in theories of overconfidence (e.g., Gigerenzer, Hoffrage, & Kleinbolting, 1991; Griffin & Tversky, 1992; for a review, see Griffin & Brenner, 2004). Gigerenzer et al. (1991) described reliance on heuristics as a basis for response

confidence on forced-choice tests of general knowledge. They argued that heuristics are typically an ecologically valid basis for confidence, and explain overconfidence on *deceptive items* as being due to nonrepresentative sampling. Our use of incongruent context corresponds to employing deceptive items. Context is typically congruent with what is said, and thus, basing confidence on the use of a fluency heuristic that reflects effects of context is ecologically valid. However, incongruent contexts do occur in the naturalistic setting and can result in real-world false hearing, like that described in the introduction. Moreover, exploration of such contexts is theoretically useful for illuminating the processes underlying age differences in hearing.

Our results and arguments should not be taken to mean that older adults should be discouraged from reliance on context for listening. Clearly, such reliance does serve the important function of allowing a means of compensation for age-related hearing deficits. Rather, it is important to highlight the potential cost of such reliance—increased probability of false hearing. In this study, hearing for words presented out of context was equated for young and older adults, and it was emphasized that context would often be misleading. Given the warning about misleading context, young adults likely engaged in careful listening, relying heavily on the sensory signal for responding and confidence, but older adults did not do so. Future work will examine situations that can encourage careful listening of the sort to avoid false hearing in older adults.

Conclusions and Future Directions

Prior investigations of age differences in audition largely ignored differences in subjective experience and have instead focused on differences in response accuracy that reflect auditory decline resulting from aging. Prior work (e.g., Humes et al., 1994; Schneider, Daneman, Murphy, & See, 2000; Schneider, Daneman, & Pichora-Fuller, 2002) has found that sensory factors are more important than cognitive factors for explaining age-related decline in speech perception. In contrast, for false hearing, large age differences in meta-audition and identification were found in the present experiments after controlling for age-related changes in speech perception ability. The convergence of age differences in false hearing with age differences in false memory and false seeing suggests a critical role for cognitive control, which may be related to age differences in frontal-lobe function (West, 1996); metacognitive performance has also been found to be quite poor among participants with frontal lobe damage (Shimamura, Janowsky, & Squire, 1990). Prior work has shown that participants with frontal lesions have a deficit in the ability to constrain responses to a veridical source when a highly accessible but incorrect alternative is present (Burgess & Shallice, 1996). These are the situations in which age group differences in false memory, false seeing, and false hearing have been observed (Jacoby et al., 2005a; Jacoby et al., 2011; Kelley & Sahakyan, 2003; Norman & Schacter, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998). An important goal for future research is to relate older adults' variability in false hearing to variability in false seeing and false remembering. A finding that participants who are more likely to falsely hear are also more likely to falsely remember and falsely see would point toward a mediating role of frontal lobe functioning.

False hearing might be sensitive to early onset of Alzheimer's disease (AD). Memory deficits in AD patients reflect decline in frontal lobe function as well as hippocampal changes (Morris, Ernesto, Schafer, & Coats, 1997). Recent work has demonstrated the utility of placing recollection in opposition to familiarity for early discrimination of healthy aging and very mild dementia of the Alzheimer's type (Tse, Balota, Moynan, Duchek, & Jacoby, 2010). False hearing may provide another valuable opposition paradigm that can be used to detect AD. False hearing may even prove to be a more sensitive measure than false memory, given the high rate of dramatic false hearing found in Experiment 2 after controlling for age-related changes in sensory decline.

There are potentially important practical implications of age differences in false hearing. Despite strong advances in the technology used in hearing aids over the past 20 years, subjective satisfaction with hearing aids has not increased (Kochkin, 2003). Hearing aids serve only to amplify aspects of the sensory signal, which is of limited utility for hearing problems, such as false hearing, that result from overreliance on context. One of the authors' (M.S.) other experiments concerns training the use of newly fitted hearing aids (Barcroft et al., 2011). When asked about the effectiveness of training, a participant in that study responded that he had not realized how bad his hearing was prior to training and use of the hearing aid. Such increased awareness of hearing deficit may serve to further increase hearing aid satisfaction and usage.

Finally, research in meta-audition could prove to be a powerful extension of existing metacognitive research. In these studies, monitoring and control frameworks of metamemory (Nelson & Narens, 1990; Koriat & Goldsmith, 1996) are extended to spoken word perception. These processes may be particularly well-suited for examination in perceptual tasks. A major advantage of meta-audition over metamemory is greater control over the physical stimuli used in laboratory experiments. To adjust identification performance in a listening-in-noise task, one needs only to increase or decrease the SNR in which the stimuli are presented. In memory, relatively complicated procedures must be employed to achieve similar ends. For example, to adjust for age differences in spoken word identification in these experiments the SNR for each participant was merely adjusted to corresponded to their SRT. In comparison, studies that attempt to eliminate age differences in memory for a control condition do so by placing young adults under divided attention at encoding, or by giving older adults additional opportunity to study (Jacoby, 1999; Kelley & Sahakyan, 2003). For applied purposes and for purposes of theory, meta-audition is as important as is metamemory, and arguably even more important. Distortions of the present, such as false hearing, may ensure distortions of the past.

References

Adobe Systems (2004). Adobe Audition (Version 1.5) [Computer software]. San Jose, CA.

American Speech–Language–Hearing Association (ASHA) (1988). *Determining Threshold Level for Speech* [Guidelines]. doi:10.1044/policy.GL1988–00008

Barcroft, J., Tye-Murray, N., Sommers, M., Mauzé, E., Schroy, C., & Spehar, B. (2011). Tailoring auditory training to patient needs with single and multiple talkers: Transfer-appropriate gains on a four-choice discrimination test. *International Journal of Audiology, 50*(11), 802–808. doi:10.3109/14992027.2011.599868

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64,* 123–152. doi:10.1037/h0043805

Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia, 34,* 263–272. doi:10.1016/0028-3932(95)00104-2

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *Journal of the Acoustical Society of America, 107,* 538–546. doi:10.1121/1.428322

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition: A Textbook for Cognitive, Educational, Life-Span and Applied Psychology.* Beverly Hills, CA: SAGE.

Gigerenzer, G., & Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528. doi:10.1037/0033-295X.98.4.506

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross-classifications. *Journal of the American Statistical Association, 49,* 732–764. doi:10.2307/2281536

Griffin, D., & Brenner, L. (2004). Perspectives on probability judgment calibration. *Blackwell handbook of judgment and decision making.* (pp. 177–198). Malden, MA: Blackwell Publishing. doi:10.1002/9780470752937.ch9

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435. doi:10.1016/0010-0285(92)90013-R

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60041-9

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging, 14,* 122–134. doi:10.1037/0882-7974.14.1.122

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D., & Lee, L. W. (1994). Factors associated with individual differences on clinical measures of speech recognition in the elderly. *Journal of Speech and Hearing Research, 37,* 465–474.

Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journals of Gerontology, 44,* P36-PP44. doi:10.1093/geronj/44.2.P36

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005a). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General, 134,* 131–148. doi:10.1037/0096-3445.134.2.131

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340. doi:10.1037/0096-3445.110.3.306

Jacoby, L. L., & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science, 15,* 49–53. doi:10.1111/j.0963-7214.2006.00405.x

Jacoby, L. L., Rogers, C. S., Bishara, A. J., & Shimizu, Y. (2011, Oct 24). *Mistaking the recent past for the present: False seeing in older adults.* Advance online publication, doi:10.1037/20025924

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005b). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language, 52,* 493–504. doi:10.1016/j.jml.2005.01.007

Jacoby, L. L., Wahlheim, C. N., Rhodes, M. G., Daniels, K. A., & Rogers., C. S. (2010). Learning to diminish the effects of proactive interference: Reducing false memory for young and older adults. *Memory & Cognition, 38,* 820–829. doi:10.3758/MC.38.6.820

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 3–22. doi:10.1037/0278-7393.25.1.3

Kelley, C.M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48,* 704–721, doi:10.1016/S0749-596X(02)00504-1

Keselman, H. J., & Rogan, J. C. (1980). Repeated measures F tests and psychophysiological research: Controlling the number of false positives. *Psychophysiology, 17,* 499–503. doi:10.1111/j.1469-8986.1980.tb00190.x

Keselman, H. J. (1998). Testing treatment effects in repeated measures designs: An update for psychophysiological researchers. *Psychophysiology, 35,* 470–478. doi:10.1111/1469-8986.3540470

Kochkin, S. (2003). MarkeTrak VI: On the issue of value: Hearing aid benefit, price, satisfaction, and repurchase rates. *The Hearing Review, 10,* 12.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517. doi:10.1037/0033-295X.103.3.490

Lovelace, E. A., & Marsh, G. R. (1985). Prediction and evaluation of memory performance by young and old adults. *Journal of Gerontology, 40,* 192–197. doi:10.1093/geronj/40.2.192

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing, 19,* 1–36. doi:10.1097/00003446-199802000-00001

Luce, P. A., Goldinger, S. D., Auer, E. T., Jr., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62,* 615–625. doi:10.3758/BF03212113

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology, 58*A, 22–33. doi:10.1080/02724980443000151

Morris, J. C., Ernesto, C., Schafer, K., & Coats, M. (1997). Clinical dementia rating training and reliability in multicenter studies: The Alzheimer's disease cooperative study experience. *Neurology, 48,* 1508–1510.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press. doi:10.1016?S0079–7421(08)60053-5

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133. doi:10.1037/0033-2909.95.1.109

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *Journal of the Acoustical Society of America, 87,* 2705–2715. doi:10.1121/1.399061

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25,* 838–848. doi:10.3758/BF03211328

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97,* 316–333. doi:10.1037/0033-2909.97.2.316

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults: One deficit or two? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46*A, 119–135. doi:1.1080/14640749308401069

Pichora-Fuller, K. (2008). Use of supportive context younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47,* S72–S82. doi:10.1080/14992020802307404

Pichora-Fuller, M., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97,* 593–608. doi:10.1121/1.412282

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275–280. doi:10.1037/h0027768

Schneider, B. A., Daneman, M., Murphy, D. R., & See, S. K. (2000).

Listening to discourse in distracting settings: The effects of aging. *Psychology and Aging, 15,* 110–125. doi:10.1037/0882-7974.15.1.110

Schneider, B. A., Daneman, M., & Pichora-Fuller, M. (2002). Listening in aging adults: From discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology, 56,* 139–152. doi:10.1037/h0087392

Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia, 28,* 803–813. doi:10.1016/0028-3932(90)90004-8

Shipley, W. C. (1967). *Shipley Institute of Living Scale.* Los Angeles, CA: Western Psychological Services.

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging, 14,* 458–472. doi:10.1037/0882-7974.14.3.458

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging, 11,* 333–341. doi:10.1037/0882-7974.11.2.333

Sommers, M. S. (2000). *Washington University Neighborhood Database.* [database]. Retrieved from http://neighborhoodsearch.wustl.edu/neighborhood/Home.asp

Tse, C. S., Balota, D. A., Moynan, S. C., Duchek, J. M., & Jacoby, L. L. (2010). The utility of placing recollection in opposition to familiarity in early discrimination of healthy aging and very mild dementia of the Alzheimer's type. *Neuropsychology, 24,* 49–67. doi:10.1037/a0014887

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13,* 230–241. doi:10.1037/0882-7974.13.2.230

Tyler, R., Preece, J., & Tye-Murray, N. (1986). *The Iowa phoneme and sentence tests,* IA: University of Iowa.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292. doi:10.1037/0033-2909.120.2.272

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59–85. doi:10.1016/0010-0285(70)90005-8

Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science, 14,* 144–148. doi:10.1111/j.0963-7214.2005.00356.x

Received April 18, 2011
Revision received September 30, 2011
Accepted October 5, 2011 ■