Summer 8-1-2013

Frequent False Hearing by Older Adults: The Effects of Predictive Context in Speech Perception

John Ryan Morton
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Frequent False Hearing by Older Adults: The Effects of Predictive Context in Speech Perception

by

John Morton

A thesis presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

August 2013

St. Louis, Missouri

Table of Contents

Acknowledgments

The author wishes to thank Mitchell Sommers for his guidance, help in revising this thesis, and friendship throughout this research manuscript and my time at Washington University. The author also wishes to thank Karl Wassmann who has been invaluable in collecting and analyzing data. The author wishes to also thank Ian Dobbins and Sandra Hale for their assistance as committee members on this thesis.

ABSTRACT OF THE THESIS

Frequent False Hearing by Older Adults: The Effects of Predictive Context in Speech Perception

by

John Morton

Master of Arts in Psychology

Washington University in St. Louis, 2013

Professor Mitchell Sommers, Chair

The current experiment investigates age differences in the subjective experience of hearing. Specifically, the experiment was conducted to examine whether older and young adults differentially weight information in the acoustic signal versus semantic context as a basis for identifying sentence-final words. Following a calibration phase, during which signal-to-babble ratios (SNBs) were individually determined to produce approximately 50% correct identification in a baseline condition, both groups were tested on word identification for sentence-final target words. The final (target) item was presented in a background babble, with the prior predictive/non-predictive semantic context presented in the clear. Three different conditions were tested using a completely within-participants design. In one condition, target words were congruent with the preceding context. In the baseline condition no semantic context was presented and in the incongruent condition, the target item was a semantically acceptable phonological competitor of the congruent trial. Results demonstrated that older adults are more susceptible to falsely hearing the contextually congruent target word on incongruent trials, with a high subjective confidence, (referred to as "false hearing") than younger adults. These results are obtained despite the fact that baseline trials produced equivalent identification performance between age groups. Different rates of false hearing, between age groups, reflect older adults' increased reliance on contextual information, rather than the acoustic signal, as a basis for responding. The greater reliance on contextual information by older compared to younger adults can be viewed as an inflexibility to adjust response techniques (sensory or context based) according to the prevailing situation. Measures of confidence and control are used to examine the degree to which the subjective hearing experience is accurate. These findings are discussed within the framework of age differences in cognitive control.

Introduction

Differences between older and younger adults' spoken word recognition have been well documented (Pichora-Fuller, Schneider, & Danemen, 1995; Sommers & Danielson, 1999), with the general finding that older adults' performance is poorer than younger adults. This finding is not surprising in light of age-related hearing loss (presbycusis); (Gates & Mills, 2005). What is somewhat unexpected, however, is that, when a semantically meaningful context is provided, age differences in spoken word recognition can be attenuated or eliminated (Sommers & Danielson, 1999; Pichora-Fuller, 1995). That is, both older and younger adults are more accurate identifying a word (e.g., shark) if it is presented in a meaningful context (e.g., "I was attacked by a ____") than if the identical signal is presented in isolation (e.g., just the word "shark"), but older adults show greater improvements than younger adults when a meaningful semantic context is provided. For example, Sommers and Danielson (1999) found that age differences in spoken word recognition were 7.15% when words were presented in isolation but 1.8% when they were presented with meaningful semantic context. This pattern of findings has been used to suggest that older adults are more proficient, than younger adults, at using context to improve word recognition. The purpose of the present work is to examine another factor that might explain the differential benefit of semantic context for spoken word recognition in older adults; namely, we want to establish whether older adults weight the contextual information more heavily than young adults, even when doing so impairs performance.

Theoretical accounts for age differences in the benefits of semantic context

One possible explanation for the differential benefit of context for older, compared to younger adults, is that demands on inhibition are decreased when semantic context is provided

(Sommers & Danielson, 1999), thereby making identification differentially easier for older adults. This inhibition account relies on the assumption that during the performance of any mental task, an individual goes through alternating states of distraction and attention, and thus requires some cognitive effort to reduce distraction, which is referred to as inhibition. Age differences in inhibition have been investigated in great detail, with the older adults generally showing poorer inhibition (Hasher & Zacks, 1988). Under the Sommers and Danielson theoretical account, when a listener is presented with a single word, or a sentence that provides no clear semantic information, listeners will have to inhibit more phonological competitors than when context is predictive. For example, if presented with the sentence "Mr. White was thinking about the cat ", the listener must inhibit a large number of phonological competitors, such as 'hat', 'cap', 'pat', 'mat', etc. However, when the listener is presented with a predictive context, for example, "The dog is chasing the cat", only items that are semantically consistent with the sentence context will be activated and will therefore require inhibition. Thus, for example, the word ‘mat’ will be activated when the word is presented in isolation, but will not be activated in the preceding sentence because it is semantically incongruent with the preceding context. That is, in the case of meaningful semantic context, inhibitory demands for suppressing activation on related items is significantly reduced because only competitors that are also semantically consistent with the context are activated.

An alternative, but not mutually exclusive, account for the reduced age differences in identifying spoken word presented in semantically meaningful compared with anomalous contexts (or in isolation) comes from the area of metacognition. Under this proposal declines in older adults' spoken word recognition are a consequence of age differences in the basis of responding. Two bases for responding are available when a listener is presented with a semantically meaningful context: the sensory signal and the available context (e.g., Nittrouer & Boothroyd, 1990). Sensation refers to the acoustic signal that reaches the ear, and is transformed into meaningful information by the brain. Context refers to the associations one builds between a given stimulus and its environment. Thus, as an older adult becomes more reliant on context to guide his or her subjective listening experience the use of the actual sensory information as a basis for responding decreases. The change in relative importance of sensory and contextual sources of information creates a bias component in which older adults use contextual information more than sensory information as a basis for identifying spoken words. According to this account, when older adults are given predictive semantic contexts (e.g., "the plumber fixed a sink") there is a greater benefit to spoken word recognition, compared to younger adults, because congruent context contains both sensory and bias components that converge on the same response. The issue regarding use of context as a basis for responding is that not all listening situations will provide a meaningful semantic context, and may sometimes contain no context whatsoever. When no context or a non-predictive context is given older adults demonstrate poorer word recognition, because they are unable to apply a sensory based response technique, instead relying on context as a basis of responding. Of critical importance is opposition procedure that would test this alternative theoretical account. A procedure of this nature must include situations in which context is predictive as well as including situations where context is non-predictive, but the target word is phonetically similar to the contextually predictive neighbor. For example, the sentence "I was attacked by a lark" provides context that is non-predictive and one phoneme different from the contextually predicted final word, "shark".

Prior studies on age differences in the basis of responding

In order to test this alternative theoretical explanation for age differences in benefits of semantic context it is necessary to have conditions in which predictive context is not congruent with the target word (e.g., “I was attacked by a lark”). Inclusion of an incongruent context condition was motivated by previous studies of memory that used an opposition procedure to dissociate recollection and familiarity as distinct bases for recall or recognition (Jacoby, et. al., 2005). Analogously, the goal of the current work is to establish whether the differential benefit of semantic context for older adults arises from differences in the use of semantic context versus the sensory signal as a basis for responding. When the context of a sentence is predictive (e.g., "I was attacked by a ____"), the listener should activate a word that is congruent with the sentence's context (e.g., shark), but if the actual word produced by the talker is phonetically similar (e.g., lark) listeners could “hear” the spoken word as the contextually congruent word. We refer to instances in which an individual identifies a word incorrectly as the contextually congruent response (i.e., they say shark when lark was actually presented) and does so with high confidence as false hearing. False hearing is especially prominent when listeners are in an environment where the signal-to-noise ratio is decreased (e.g., crowded public area such as a mall). Using contextually incongruent stimuli researchers can then test if older adults are in fact using a different basis for responding. For example, perhaps false hearing would be more prevalent in older adults due to a loss of flexibility regarding response techniques; such that, when an older adult is in a situation where it would be advantageous to make perceptual judgments on the basis of sensory information, they are unable to make this adjustment. For example, when an older adult is presented with the sentence, “I was attacked by a lark”, the older adult would benefit from using sensory information as the basis for responding. Using sensory information in this situation may or may not lead the individual to correctly hear the target word (eg., the older adult might hear “lamp” instead of “lark”), but using sensory information as a basis of responding would not likely lead to false hearing. This is to say that contextual information is not required when using sensory information as a basis for responding, but false hearing is contingent upon it. However, as the older adult is unable to make the adjustment to using sensory information as a basis for responding, he or she will continue to rely on contextual information, which leads the individual to falsely hear the word “shark”.

Recent evidence from both cognitive (memory) and perceptual (seeing and hearing) provides strong evidence to suggest age differences in the bases for responding to items in context; demonstrating that older adults are more susceptible to both false seeing (Jacoby et al., 2011) and false hearing (Rogers et al., 2012).

To support the theory of a differential bases for responding between age groups Jacoby et. al (2011) investigated age differences in false seeing occurrences. In this experiment participants were, after learning the paired associates, visually presented with a prime word prior to a very brief presentation of a target word that included both a forward and backward visual mask. On a congruent trial, the prime (e.g., DART) was the same word as the target. On incongruent trials the target word was an orthographic neighbor of the prime word differing only in a single letter (e.g., DIRT). After viewing the prime and target words participants were given a fragment completion task in which the goal was to fill in the missing letter he or she believed was the target word (e.g., d_rt). If older adults have a different basis for responding, one would expect large age differences in incongruent trials where older adults completed the word fragment with the primed context (e.g., the letter “a” for prime “dart”). Jacoby et. al. (2011) reported that older adults were significantly more likely to reproduce fragment completions for the primed word in incongruent trials than were younger adults. More compelling evidence for the differential use of context based responding came from trials in which participants were not presented a target word at all (e.g., just a blank slide). On these trials an individual should simply be guessing what the target word was because no target was actually provided. Under these conditions older adults reported seeing the primed word in the blank screen on 34% of the trials, whereas younger adults reported the primed word only 8% of the time.

In a study paralleling the false seeing data (Jacoby et al., 2011), Rogers et. al. (2012) created an experiment to test age differences in false hearing using a paired associate procedure. Participants were first instructed to learn semantically related word pairs (e.g., BARN-HAY). Once participants were able to recall all of the pairs (as assessed by providing the first item and asking for the second), they were then presented with the first item (e.g., BARN) in the clear followed by either the learned associate (e.g., HAY), or a phonologically related item (PAY), with this second item being presented in a background babble that was set individually for each participant in order to equate audibility. In the case of contextually congruent word pairs (e.g., BARN-HAY) both younger and older adults should demonstrate high levels of confidence and word recognition, because both bases of responding (associative context and sensory signal) converge on the same response (e.g., HAY). Note that in previous studies of age differences in the benefits of context, this congruent condition, along with a baseline measure assessing performance without context, are the only ones that have been presented, and therefore do not allow assessment of differences in the bases of responding. The use of an incongruent condition, however, allows exactly this type of dissociation. Specifically, if older and younger adults differ in their bases of responding, older adults should exhibit not only more incorrect responses on incongruent trials – where they respond ‘HAY’ to the initial item ‘BARN’, even though ‘PAY’ was actually presented – but they should provide very high confidence measures on those incorrect responses. Consistent with this prediction, Rogers et. al. (2012) reported that older adults were significantly more likely than younger adults to respond incorrectly on incongruent trials, and confidence on those incorrect responses was significantly higher for older compared to younger adults. These findings are consistent with the proposal that older adults, compared to younger adults, are more reliant on context as a basis for responding, and that they are having the subjective hearing experience of hearing the congruent item on incongruent trials.

Present experiment

Although previous studies have shown that there are differences in false hearing (and false seeing) rates between younger and older adults these studies (Jacoby et al., 2011; Rogers et al., 2012) have relied exclusively on the paired-associate paradigm. In the present study, we wanted to establish whether age differences in false hearing would increase with the use of more ecologically valid stimuli, in this case sentences in which a preceding context either was, or wasn't predictive, of a sentence-final word. This is to say that it is entirely possible that by using sentences, the additional context would increase false hearing rates far more than in paired associate paradigms. Therefore, in the present study, identification of sentence-final items was measured using materials from the speech-perception-in-noise (SPIN) test (Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984). Participants were presented with sentences in which all but the final word was presented in the clear, and the target (sentence-final) item was presented at individually adjusted signal-to-noise ratio (to equate audibility differences between individuals and age groups). The present study used three different contextual conditions: 1) congruent - the sentence final word was highly predictable from the preceding context ; 2) baseline - the sentence final word was not predictable from the preceding context; 3) incongruent - the sentence final word completed a meaningful sentence, but was not the item most predictive by the context and was phonologically similar to the target-final item in the congruent condition. On congruent trials, we used unmodified versions of the SPIN high-predictability sentences (e.g., "The shepherd watched his sheep"). On baseline trials, participants heard low-predictability versions of the SPIN sentences (e.g., "Paul heard they asked about the rice"). Low predictability trials therefore contained little to no contextual information, and, consequently, correct responding could only be done on the basis of sensory information. Finally, on incongruent trials we presented modified versions of the high-predictability sentences in which the target item was replaced by a plausible, but less predictable, phonological neighbor (e.g., "The shepherd watched his sheath").

To examine age differences in the willingness to act on perceptual experience, following each identification response, participants were asked whether they would like to "volunteer" or "withhold" their identification response (Koriat & Goldsmith, 1996), and were also asked to rate their confidence (0 to 100%) as to how sure he or she was that the word they reported was the word they actually heard in noise.

We hypothesized that older adults would be more likely than young adults to provide an incorrect, but context consistent, response in the incongruent condition (i.e., responding "shark" when provided the sentence "I was attacked by a lark"). This hypothesis is based on findings that older adults are more likely than younger adults to use context, rather than sensory information, as a basis of responding (Jacoby et al., 2011; Rogers et al., 2012). It was also expected that both age groups would produce the highest correct identifications on congruent trials, but that older adults would actually perform better than young adults in this condition because both sensory and contextual information converge on the correct target item. We also predict that false alarms on incongruent trials (false alarms are defined here as trials in which the participant responds with the contextually predicted item despite the fact that the phonological neighbor was actually presented) will be significantly greater for older adults compared to younger adults. We hypothesize that, within these incongruent false alarms, older adults will have significantly higher confidence in their responses than younger adults.

Methods

Participants

A total of twenty eight undergraduate students were recruited through the Washington University subject pool, and received either ten dollars or one course credit per hour. Participants were calibrated at the start of the experiment to a signal-to-babble ratio that produced an accuracy of 50% on baseline trials. Occasionally this ratio was set either too high or too low resulting in values approaching ceiling or floor effects respectively (participant scores above 65% or below 35% on baseline trials). Of these twenty eight younger adults four were excluded due to poor calibration, leaving a total of twenty four participants. These young adults ranged in age from 18 to 29 years (M = 22.46, SD = 3.83). A total of thirty five older adults were recruited through the Washington University Older Adult subject pool, and received ten dollars per hour of participation. Of these thirty five older adults eleven were excluded due to poor calibration, leaving a total of twenty four participants. These older adults ranged in age from 66 to 78 years (M = 72.88, SD = 5.161). Participants were tested on the Vocabulary subset of the Shipley Institute of Living Scale (Shipley, 1967). The mean score was slightly higher for older adults (M = 35.8, SD = 3.94) compared to younger adults (M = 35, SD = 3.36), but this was not statistically significant, t(46) < 1, *ns*. None of the older participants reported taking medications or having health conditions that would affect cognitive function. All participants reported normal or corrected-to-normal vision.

Materials

A critical issue for any study comparing speech perception in older and young adults is that older adults typically have reduced audibility of spoken words due to presbycusic hearing loss. In order to control for individual differences in hearing we used a modified version of the ASHA speech reception threshold (SRT) titration procedure (ASHA, 1998). In this procedure signal-to-noise ratio (SNR) was varied adaptively to determine the SNR value that produced 50% correct identification for words. A set of 60 isolated words was used as stimuli for the titration procedure.

A total of 126 sentences were taken or modified from the SPIN sentence test (Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984), and used for the study. Six of these sentences (two sentences per trial type) were used for practice prior to beginning experimental trials in order to ensure that each participant had a sufficient understanding of the task at hand. Of the remaining 120 sentences, 40 each were presented in congruent, baseline, and incongruent conditions. The congruent condition used unmodified versions of high-predictability SPIN sentences in which the final word was highly predictable from the prior semantic context (The shepherd watched his sheep). The baseline condition used unmodified versions of low-predictability SPIN sentences, in which the context provides minimal information about the sentence-final item (e.g., Paul heard they asked about the rice). Finally, the incongruent condition used modified versions of the SPIN sentences in which the final (target) word was replaced by a phonological competitor confusable with the original target item but that still produced a meaningful sentence (e.g., the shepherd watched his sheath). Half of these replacement items differed from the original target in the initial phoneme (i.e., take the sentence, ‘I was attacked by a shark’, and replace the original target item with ‘lark’) and half had the final phoneme changed (i.e., take the sentence, ‘the shepherd watched his sheep’, and replace the original target item with ‘sheath’). Across the three conditions, final target items were equated for frequency and neighborhood density. In addition, for the congruent and incongruent conditions, presentation of intact or modified high-predictability sentences was counterbalanced across participants such that an equal number received each version of a sentence (i.e., half heard the ‘sheep’ ending for the sentence context “the shepherd watched his _______” and half received the ‘sheath’ ending). No participant received both versions of any of the sentences.

All of the auditory stimuli were spoken versions of the above sentences recorded at 48,000 Hz with a 16-bit resolution, using a table-mounted microphone (Shure PG27) in a double-walled sound attenuating booth. Sentences were spoken by a male speaker with a standard American dialect. Root-mean-square (RMS) amplitude of the stimuli was equated. Stimuli were then down sampled to 11,025 Hz using Adobe Audition for presentation in the experiment.

The auditory stimuli were masked (full word in calibration phase and final word in sentence tests) using a 6-talker babble. The babble was captured from the Iowa Auditory visual Speech Perception Laserdisc (Tyler, Preece, & Tye-Murray, 1986) using a 16-bit converter at a sampling rate of 44,100Hz and then down sampled to 11,025 Hz using Adobe Audition. A different random sample of the babble was used for each presentation.

Procedure

Calibration Phase

The procedure in the experiment was broken into two phases: the calibration phase and the perception test phase. The purpose of the calibration phase was to find a signal-to-noise ratio that produced 50% accuracy for word identification using a modification of the American Speech-Language-Hearing Association's recommended procedure (ASHA, 1998) for obtaining a Speech Reception Threshold. That is, the goal of the calibration phase was to equate audibility by determining the SNR that would produce 50% correct identification of individual words. During the calibration phase participants were seated in a testing room, and stimuli were presented binaurally over headphones (Beyerdaynamic DT 801). Participants were told to repeat the presented word back to the experimenter, and that, if he or she had no idea what the presented word was, to give the best possible guess. For each correctly identified word the signal-to-noise ratio was decreased by 2dB SPL, and for each incorrectly identified word the signal-to-noise ratio was increased by 2dB SPL. Calibration was completed when participants responses stabilized such that if the signal-to-noise ratio was decreased he or she would incorrectly identify the word, but when the signal-to-noise ratio was increased he or she would correctly identify the word. After phase completion the average dB level was obtained to be used in the perception test phase. Stimuli were randomly presented from the total calibration word list.

Perception Test Phase

During the 120-trial perception test phase, there were three different trial types: congruent, incongruent, and baseline trials. Trials of each type were presented in a single block with the restriction that no more than three trials of a given type were presented consecutively. Participants were seated facing the computer screen in the same testing room used for the calibration phase. Participants were informed that they would be hearing a series of sentences with the final word in a babble-background noise. Participants were told that the masking noise would sound like when you first walk into a room with several people talking at once. Participants were told that their task would be to repeat the final word of this sentence back to the experimenter, and that if they had no idea what this word was to simply respond with the best possible guess. Participants were warned that there would be three types of sentences: predictive sentences like, "the plumber fixed the <u>sink</u>", misleading sentences like, "the plumber fixed a <u>drink</u>", and sentences that would have no predictive qualities at all like, "Mr. White is thinking about the <u>drink</u>". Because of this, participants were instructed to respond on the basis of what they heard in noise, not what the context of the sentence may or may not lead them to believe.

After participants provided the target word identification they were instructed to indicate how confident they were that the response provided was, in fact, the word presented in noise. Participants gave this confidence rating using a 0-100 percent scale. Participants rated aloud, and were encouraged to use the full range of the scale. As with the identification judgments, participants were instructed to make their confidence judgments only on the basis of what they heard in noise.

To investigate any possible age differences in the willingness to act on their perceptual experiences, after providing a 0-100% confidence rating, participants were asked to volunteer or withhold this response for scoring (Koriat & Goldsmith, 1996). Participants provided this response by speaking aloud if he or she would like to volunteer or withhold the trial. Participants were informed that volunteering a correctly identified target word would result in a point added to his or her total score. If the participant volunteered and incorrectly identified the target word he or she would have a point deducted from his or her total score, and, if he or she withheld the response, there would be no change to the total score.

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 (0-100), (c) the decision to volunteer or withhold the response for scoring, and (d) the potential misleading nature of sentence context. The experimenter verbally repeated instructions and questioned participants until the participants' procedure report was complete. All participants' procedure reports were complete before the beginning of the perception test phase.

The timing for each trial was as follows: 200 ms before the sentence was presented over the headphones a single asterisk "*" was presented visually in the top center portion of the screen until the onset of the aurally presented sentence. The babble-background noise began approximately 30 ms before the target item and ended with the offset of the target word. Participants were given no time limit with regards to reporting the target word, confidence, or the decision to volunteer or withhold the response.

Results and Discussion

Correct Identifications

Unless noted, only effects found significant with an alpha less than .05 are reported. All of the excluded participants who were eliminated due to poor calibration were included in the analysis to ensure findings did not significantly differ from those obtained when these participants were excluded. All significant and non-significant findings were unaltered by the inclusion of these participants, and, as such, they were excluded from further analyses. Identification accuracy was measured as the proportion of trials on which participants correctly identified the target word in noise. To confirm the statistical reliability of these findings, correct identifications were analyzed using a 2 (age: younger, older) X 3 (trial type: baseline, congruent, incongruent) repeated measures mixed model analysis of variance (ANOVA), revealing a significant age x trial type interaction, $F(2, 92) = 4.749$, $MSE = .007$, $p < .05$, $\eta_p^2 = .094$. Figure 1 displays the proportion of correct identifications by trial type and age group.

Consider first, the baseline condition in which context is entirely non-predictive of the sentence-final item. Performance for this condition did not differ significantly between young and older adults and overall accuracy was very close to the targeted 50% correct rate, $ts < 1$, *ns*. Thus, the baseline condition indicates that the procedure for equating audibility was successful. Next, consider results for the congruent condition in which both context and sensory information converge on the same response. Unlike some previous studies of speech perception and aging (Sommers & Danielson, 1999), older adults correctly identified significantly more words than young adults, $t(46) = 2.47$, $p < .05$. Finally, analysis of incongruent condition correct identifications did not reveal any significant difference between age groups, $ts < 1$, *ns*.

If we compare the pattern of results for the congruent correct identifications, with that of the incongruent false alarms (e.g., when the participant response is the contextually predictive phonologically competitor to the word actually presented), the pattern of accuracy reverses for older and young adults. That is, when sensory information and semantic context lead to different responses, and older adults are more likely than younger adults to respond incorrectly on the basis of context. As Figure 1 displays (right of the dividing line), older adults ($M = 0.5$, $SD = 0.2$) were significantly more likely to produce false alarms on incongruent trials as opposed to younger adults ($M = 0.39$, $SD = 0.11$) , $t(46) = 2.2$, $p < .05$.

Confidence data

To investigate confidence results a 2 (age: younger, older) X 3 (trial type: baseline, congruent, incongruent) repeated measures mixed model analysis of variance (ANOVA), revealed a significant age x trial type interaction for confidence, $F(2, 92) = 6.178$, $MSE = .139$, $p < .01$, $\eta_p^2 = .118$. The mean confidence ratings assigned to correct identifications in the congruent, baseline, and incongruent conditions are displayed in Figure 2 as well as false alarms in the incongruent conditions. Consistent with older and young adults being restricted to sensory information as a basis for responding in the baseline condition, confidence did not differ as a function of age, ts < 1, *ns.* In contrast, for the congruent condition, confidence was higher for older than for young adults, $t(46) = 2.05$, $p < .05$. Analysis revealed that older adults ($M = 50.33$, $SD = 13.46$) were numerically less confident in their incongruent correct identifications compared to younger adults ($M = 57.55$, $SD = 13.3$), but this did not reach significance, ts < 1, *ns*.

Confidence in incongruent false alarms is of greatest importance in the present experiment as they speak to age group differences in context use. Confidence in incongruent false alarms was expected to differ significantly between older and younger adults, but analysis actually showed no significant differences, ts < 0.2, *ns*. Although the data trended in a manner similar to previous studies (Rogers et al., 2012), in which older adults mean confidence in incongruent false alarms was significantly higher than younger adults, we did not obtain a significant age difference with respect to confidence in the false alarms.

Dramatic False Hearing

To further examine age differences in the basis of responding, we conducted an additional analysis of the incorrect responses favored by context in the incongruent condition. Specifically, we examined what we have termed dramatic false hearing – defined as providing the (incorrect) response favored by context with a confidence rating of 100% (i.e., cases in which individuals were 100% confident in their incorrect responses). Older adults (M = 0.16, SD = 0.22) demonstrated significantly more occurrences of dramatic false hearing compared to younger adults (M = 0.04, SD = 0.1), $t(46) = 2.337$, $p < .05$.

Resolution

Resolution is a measure of metacognitive monitoring that assesses the extent to which confidence in a response is related to accuracy. Resolution was measured in two ways; using gamma correlations and confidence discriminability. Gamma correlations were used at the item-level to examine the correspondence between confidence and accuracy (Goodman & Kruskal, 1954). Gamma correlations range from +1 to -1, where a strong positive value on this scale refers to a confidence judgment that distinguishes well between correct versus incorrect responses. We expected older adults' resolution should be quite good in congruent trials, moderate in baseline trials, and poor in incongruent trials. We predicted younger adults, however, would demonstrate similar gamma correlations in baseline and congruent conditions, but, in incongruent conditions, we predicted younger adults would differ from older adults by demonstrating significantly higher gamma correlations. If so, this would indicate that older adult participants' monitoring ability is strong when he or she can use the predictive context, but, when the context offers an incongruent predictive context, their monitoring ability suffers. Since younger adults are better suited to use sensory information as a basis of judgment, they should not demonstrate this issue.

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. There were ten participants (four younger and six older adults) who produced this issue and were excluded from the analysis. Figure 3 shows the resolution data from the remaining 38 participants.

The resolution data in Figure 3 support the prediction that participants' demonstrate good monitoring in congruent trials, and poor monitoring in incongruent trials. However, only older adults produced a negative gamma correlation on the incongruent trials. The 2 (age: younger, older) X 3 (trial type: baseline, congruent, incongruent) repeated measures mixed model analysis of variance (ANOVA), did not reveal a significant age x trial type interaction for resolution, $F(2, 72) = .193$, $MSE = .074$, $p = ns$, $\eta_p^2 = .005$.

Although the interaction was not significant the data in figure 3 show that, for all subjects, resolution in the congruent condition was quite good in that, if a participant was highly confident in his or her response, it was very likely for him or her to correctly identify the target word. In the incongruent condition, resolution was noticeably reduced for both groups compared to congruent conditions, but only older adults show a negative gamma correlation in the incongruent trials. In fact older adults' gamma correlations in the incongruent trials demonstrate that, the more confident the older adult was in his or her response, the more likely he or she was to provide an incorrect response.

Another common technique for analyzing how efficient a participant is at determining whether a given response was correct of incorrect, using confidence, is confidence discriminability (Hart, 1965). Confidence discriminability takes the average confidence for correct responses and subtracts this value from the average confidence in incorrect responses. Because some participants' scores could not be used in the gamma analysis, we desired to ensure our results replicated with other measures that could include all participants' scores. The advantage of using this measure, as compared to gamma correlations, is that no participants need be excluded due to the above described reasons. The 2 (age: younger, older) X 3 (trial type: baseline, congruent, incongruent) repeated measures mixed model analysis of variance (ANOVA), revealed a significant age x trial type interaction for resolution, $F(2, 92) = 3.356$, $MSE = 474.361$, $p < .05$, $\eta_p^2 = .068$. Figure 4 shows the mean confidence discriminability scores on congruent, baseline and incongruent trials for both younger and older adults. Planned comparisons demonstrate that older adults ($M = -30.4$, $SD = 21.51$) produced significantly lower confidence discriminability than younger adults ($M = -13.78$, $SD = 13.59$) on incongruent trials, $t(46) = 3.2$, $p < .01$. There were no significant differences in confidence discriminability on congruent or baseline trials between age groups, $t(46) < 1$, *ns*. These findings differ with the gamma correlation analysis, and this makes understanding the true relationship between resolution differences between younger and older adults difficult. Both methods have disadvantages, but of greater interest is what both analyses converge on. The commonality seen in both resolution measures is that older adults have poorer resolution on incongruent trials than younger adults.

Metacognitive Control Measures

In addition to examining age differences in subjective experience of hearing, we also wanted to investigate the extent to which older and younger adults differ on their willingness to act on their subjective experiences. Thus, in the current study, participants were given an opportunity either to have the trial count toward their total score or to withhold the response from scoring. Previous studies have shown that rate of volunteered responses strongly relates to confidence in those responses (Koriat and Goldsmith, 1996; Rogers et al., 2012). Therefore, we predicted similar rates of volunteering for the congruent and baseline conditions across age groups because as noted earlier confidence did not differ as a function of age in either of these two conditions. For the incongruent condition, however, we expected that older adults would be more likely than young listeners to volunteer an incorrect false alarm because their subjective experience of hearing is based on the preceding context to a greater extent than young adults, leading to higher confidence ratings for older than for young adults in the incongruent condition.

In contrast to these predictions, in the present experiment there were no significant differences in the rates at which items were volunteered on congruent, baseline or incongruent trials between older and younger adults, all $t(46) < 1$, *ns* . Possible explanations for the absence of age differences in volunteering responses are considered in the General Discussion.

Of greater concern to the present experiment is whether there would be differences on the rates of correctly volunteered responses in congruent, baseline and incongruent trials. In other words, although differences in rates of volunteered responses between age groups and conditions is of interest, the true measure of whether participants accurately control their responses comes out of the times in which they volunteer or withhold responses correctly. If it is the case that older adults rely less on sensory information when a congruent or incongruent context is given then their rate of correctly volunteered responses should be lower in the incongruent condition, compared to younger adults. These values are obtained by taking the total number of correctly volunteered responses and dividing this value by the total number of volunteered responses. The 2 (age: younger, older) X 3 (trial type: baseline, congruent, incongruent) repeated measures mixed model analysis of variance (ANOVA), revealed a significant age x trial type interaction for resolution, $F(2, 92) = 3.118$, $MSE = .021$, $p < .05$, $\eta_p^2 = .063$. Figure 5 displays the proportion of correctly volunteered trials by condition. As predicted, younger adults ($M = 0.26$, $SD = 0.15$) are significantly more likely to correctly volunteer a response on incongruent trials compared to older adults ($M = 0.14$, $SD = 0.18$), $t(46) = 2.42$, $p < .05$. There were no significant differences between proportion of correctly volunteered items in congruent or baseline conditions between age groups, $t(46) < 1$, *ns.*

General Discussion

The present study investigated whether age affects susceptibility to false hearing using stimuli that are more ecologically valid than those used in previous studies. Results showed that on incongruent trials (when context is misleading) older adults falsely heard the contextually congruent word significantly more often than did younger adults. This study showed that older adults were more likely to correctly identify the target item on congruent trials. Confidence data revealed a significant difference in congruent items; in which older adults were more confident in their correct identifications than younger adults. No significant difference was found, between age groups, regarding confidence in incongruent false alarms (false hearing), which was unexpected.

Results showed that on incongruent trials (when context is misleading) older adults falsely heard the contextually congruent word significantly more often than did younger adults. This finding is consistent with those reported by Rogers, et. al. (2012) and Jacoby et. al.(2011) where older adults were also significantly more likely to demonstrate incongruent false alarms than younger adults. Furthermore, this study showed that older adults, compared to younger adults, were more likely to correctly identify the target item on congruent trials, as well as being significantly more confident in those responses. This finding is also consistent with the findings reported by Rogers, et. al. (2012) and Jacoby et. al (2011).

Inhibition and response bias as accounts for age differences

Taken together, these findings are consistent with the proposal that the differentially greater benefit that older adults obtain from adding a meaningful semantic context, relative to young adults, is due in part to differential bases of responding. In other words if older adults did not differ from younger adults in their basis for responding we should expect equal levels for false seeing/hearing once age differences in vision/hearing are equated. The fact that research consistently finds significant differences between younger and older adult's correctly identified congruent items demonstrates a difference between bases for responding between these age groups.

Younger adults also had more correct identifications in incongruent conditions than older adults. These findings are in accordance with previous findings showing that older adults compensate for hearing loss by increasing the use of contextual cues (Hutchinson, 1989; Pichora-Fuller, 2008; Sommers & Danielson, 1999; Wingfield et al., 2005). These findings support the theoretical claim that age differences in spoken word recognition are due to a differential basis for responding, but do not discount, at least some influence, of inhibition differences between age groups (Sommers & Danielson, 1999). The inhibition account fits well with the overall trends seen between the differing trial types. For example, the inhibition account would predict that congruent trials would produce the highest proportion of correctly identified target words, and the incongruent condition the least. This is to say that when context is predictive of the final word there would, in an inhibitory account, be fewer phonological competitors activated (for example, in the sentence "Playing checkers can be <u>fun</u>", the words 'shun', 'bun', 'gun', etc. would not be activated due to the sentence's syntax), thus correct identifications would be more probable. On incongruent trials, with context now misleading, an inhibition account would predict lowest proportion of correctly identified target words, because, although the number of competitors is the same as on a congruent trial, the context heavily favors only one competitor (the word favored by the context). The current experiment is not able to address the role of inhibition compared to the role of a differential basis for responding, and further research is needed to examine whether these theories are in fact mutually exclusive or not.

<u>Age differences in dramatic false hearing</u>

Further analysis revealed that older adults provided significantly greater numbers of dramatic false hearing occurrences on incongruent trials than younger adults. As a reminder dramatic false hearing refers to erroneously selecting the alternative favored by context in the incongruent condition and expressing 100% confidence that the selected word was the one presented in noise. If it is assumed that a confidence rating of 100% truly represents the subjective experience of hearing the contextually congruent word, despite the fact that a phonologically similar was actually presented, than dramatic false hearing occurrences are of great interest. Rogers, et. al. (2012) also reported older adults having significantly more dramatic false hearing occurrences than younger adults. These findings are also in accordance with a theory of differential bases for responding between age groups. If older adults truly rely on context as a basis for responding to a greater degree than younger adults it should be expected that this subjective experience of hearing the contextually congruent word on incongruent trials be manifested in stronger confidence in incongruent false alarms.

Age differences resolution

The present experiment examined resolution between age groups, which showed differences in participants' confidence, discriminating from whether he or she was correct, or incorrect, on a given trial. Researchers first used gamma correlations to investigate if there were differences between older and younger adults' resolution. Although the interaction did not reach significance, the analysis did provide insight into differences between age groups. As opposed to the younger adults', older adults' averaged gamma correlation was negative, indicating that the more confident an older adult was on an incongruent trial, the more likely he or she was to provide an incorrect response. A second measure of resolution was used to ensure that the gamma correlation analysis was accurate due to the large number of participants having to be excluded for either always, or never, correctly identifying the target word. The analysis of confidence discriminability did reveal a significant age by trial type interaction. Older adults demonstrated significantly lower confidence discriminability scores than younger adults on incongruent trials. Together, these resolution measures provide further support to the claim that older adults are less able to use sensory information as a basis for responding, because, when context is not predictive, their resolution greatly diminished.

Divergent results between present and past experiments

Although many results of the present study were in accordance with previous research there were a number of divergent and novel results. The present experiment produced different findings, compared to previous studies (Rogers et al., 2012), concerning confidence, resolution and rates of volunteerism. First, no significant difference between age groups was found on confidence in incongruent false alarms. Secondly, no significant age by trial type interaction was obtained when analyzing the resolution using gamma correlations. This was unexpected as an age by trial type interaction has previously been shown to have a significant interaction (Rogers et al., 2012). Finally, looking into metacognitive control measures, the present experiment showed no age differences in the rate of volunteered responses.

There are at least two possible reasons for these differences, between current and previous experiments, we believe could account for such differences. The first possibility is that confidence differed between front and back substitutions on incongruent false alarms. If, for example, age groups had equal confidence on back substitutions, but unequal confidence on front substitutions, then a significant difference between incongruent false alarms may have been diminished. This was not the case as analysis revealed no significant differences between age groups confidence on incongruent false alarms for front or back substitutions. The second possible account for the differences in confidence findings between current and past experiments is that using full sentences uncovers a far more complex relationship between confidence and stimulus type. In other words, the differences seen between this study and previous experiments might be due to the methodological differences between using paired associates compared to full sentences. The present experiment used full sentences, which offer a more robust context to the final target word, as well as longer stimulus presentation duration, than does a paired associate. This additional stimulus duration and context may alter the degree to which confidence and cognitive control measures manifest. It is also important to know that in Rogers, et. al. (2012), experimental design context was created by teaching participants the paired associates prior to testing procedures. The present study did not include any learning of the stimuli set used in the testing procedure. This methodological difference may have also contributed to the differences seen in both confidence and cognitive control measures because prior exposure may influence these measures. For example, it is possible that prior exposure makes participants more confident on incongruent false alarms, and this increased confidence directly relates to his or her willingness to volunteer a trial. Further research is needed to understand the divergence on some of the results in the present experiment compared to Rogers, et. al (2012).

Implications and limitations

Results from the present experiment and Rogers, et. al. (2012) demonstrated that older adults, compared to younger adults, are more likely to demonstrate false hearing occurrences on incongruent trials. Research has also shown that, on incongruent trials, older adults are more likely to falsely see (Jacoby et. al., 2011) and falsely remember (Hay & Jacoby, 1999) than are younger adults. This convergence in results between these different domains suggests a general deficit in cognitive control underlying these age differences, and, although each specific domain may be influenced by other factors such as deficits in hearing, vision or memory, a central process argument is further supported by the present experiment. A general deficit in cognitive control explanation for age differences becomes even strong when one considers that, in both false seeing and false hearing domains, visual acuity and audibility, respectively, participants were individually equated. Thus both younger and older adults demonstrated equal identification accuracy on baseline or control conditions.

The age differences found in false remembering (Jacoby et al., 2005a) have been interpreted in terms of a dual-process model of memory, and we propose a similar process underlies the age differences in false seeing and hearing. In the domain of memory, dual-process model posits that conscious recollection and implicit influences (familiarity) form the two possible bases for responding to a stimulus. Conscious recollection is a more effortful process where retrieval of some specific elements of a stimulus is used, whereas familiarity represents a more automatically processed and holistic basis for responding. We postulate that a similar process occurs in false hearing and false seeing. Similar to conscious recollection, when a participant responds to an item using sensory information he or she is using the specific speech signal. However, when a participant falsely hears the target word predicted by context on an incongruent trial he or she is likely not weighing sensory information as strongly as contextual information as the basis for responding. Instead, analogous to using familiarity as the basis for responding in memory retrieval, the participant is using the far more holistic context in the stimulus as a basis for responding in perceiving spoken word. Thus, it seems that younger adults are more likely to restrain their basis for responding to the sensory signal than older adults when given the knowledge that certain trials may be misleading.

The dual-process model is limited in the present experiment. In order to truly support a general deficit in cognitive control to explain age differences in false hearing, seeing and memory researchers need to test participants in all three domains. A design of this nature allows researchers to investigate the degree to which individuals incongruent false alarms correlate with incongruent false alarms in the other domains. For example, if an individual has a high degree of incongruent false alarms in a false hearing condition he or she should be likely to have a high degree of incongruent false alarms in both false seeing and remembering conditions. Such a finding would add a great deal of support to the claim that age differences are caused by a general deficit in cognitive control. Future research is needed to fully understand the degree to which cognitive control differs between the age groups, but the present study and past studies offer support for this account.

The present experiment also is limited to situations in which there is no differing motivation between conditions or trials within the conditions. Of interest is the degree to which bases for responding can be altered by motivation. For example, it is possible that if an individual were to put a high priority on correctly identifying an incongruent trial, incongruent false alarms may decrease on high priority trials compared to non-high priority trials. If this was found to be the case for older adults, it would imply that older adults are able to alter their basis for responding to the more optimal strategy (eg., using sensory information as the basis for responding to an incongruent trial) when motivation is increased. Future research is needed to understand the manner in which motivation affects the basis for responding, as in high priority trials individuals may chose to deviate from the basis response he or she may be using.

Conclusion

The present experiment provides strong evidence that older adults are less flexible concerning their basis for perception (eg., using sensory or contextual sources of information) compared to younger adults. The present experiment expands the literature on false hearing by using more ecologically valid stimuli, and finding areas of further research interest regarding confidence differences. We believe researchers will need to focus more on these ecologically valid stimuli if the field is to progress into understand the everyday human experience of speech perception, and how this changes with age.

References

American speech-language-hearing association (ASHA) (1998). *Determining Threshold Level for Speech* [Guidelines]. doi:10.1044/policy.GL1988-00008.

Bilger, R. C., & Nuetzel, J. M. Rabinowitz, & Rzeczkowski, C.(1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*, *27*, 32-48.

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet*, *366*(9491), 1111-1120.

Goodman, L. A., and Kruskal, W. H. (1954). Measures of association for cross classifications*. *Journal of the American Statistical Association*, 49(268), 732-764.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56(4), 208.

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: 10.1037/08827974.14.1.122

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.

Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journal of Gerontology*, 44(2), P36-P44.

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*, 44, P36-PP44. Doi: 10.1093/geronj/44.2.P36.

Jacoby, L. L., Rogers, C. S., Bishara, A. J., & Shimizu, Y. (2011, Oct24). *Mistaking the recent past for the present: False seeing in older adults*. Advance online publication, doi: 10.1037/20025924.

Koriat, A., and Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review; Psychological Review*, 103(3), 490.

Nittrouer, S., and Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. *The Journal of the Acoustical Society of America*, 87, 2705.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*, 593.

Pichora-Fuller, M. K., (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, 47(S2), 72-82.

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and aging*, *27*(1), 33.

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*(3), 458.

Tyler, R. S., Preece, J. P., & Tye-Murray, N. (1986). The Iowa phoneme and sentence tests. *Iowa City, The University of Iowa, Department of Otolaryngology*.

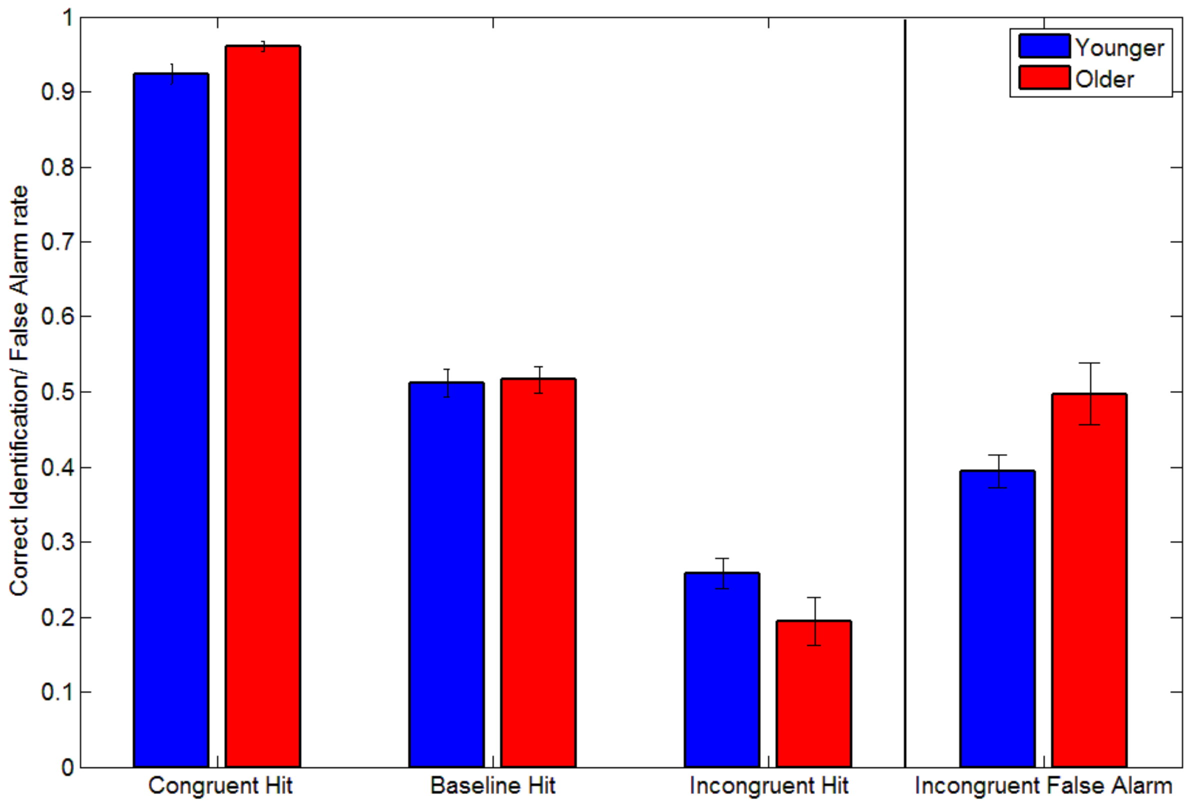

Figure 1. Proportion of correct identification's and incongruent false alarms made for each trial type. Correct identifications for congruent, baseline and incongruent conditions are plotted to the left of the dividing line. Proportion of incorrect responses, predicted by context, on incongruent trials (false alarms) are plotted to the right of the dividing line.

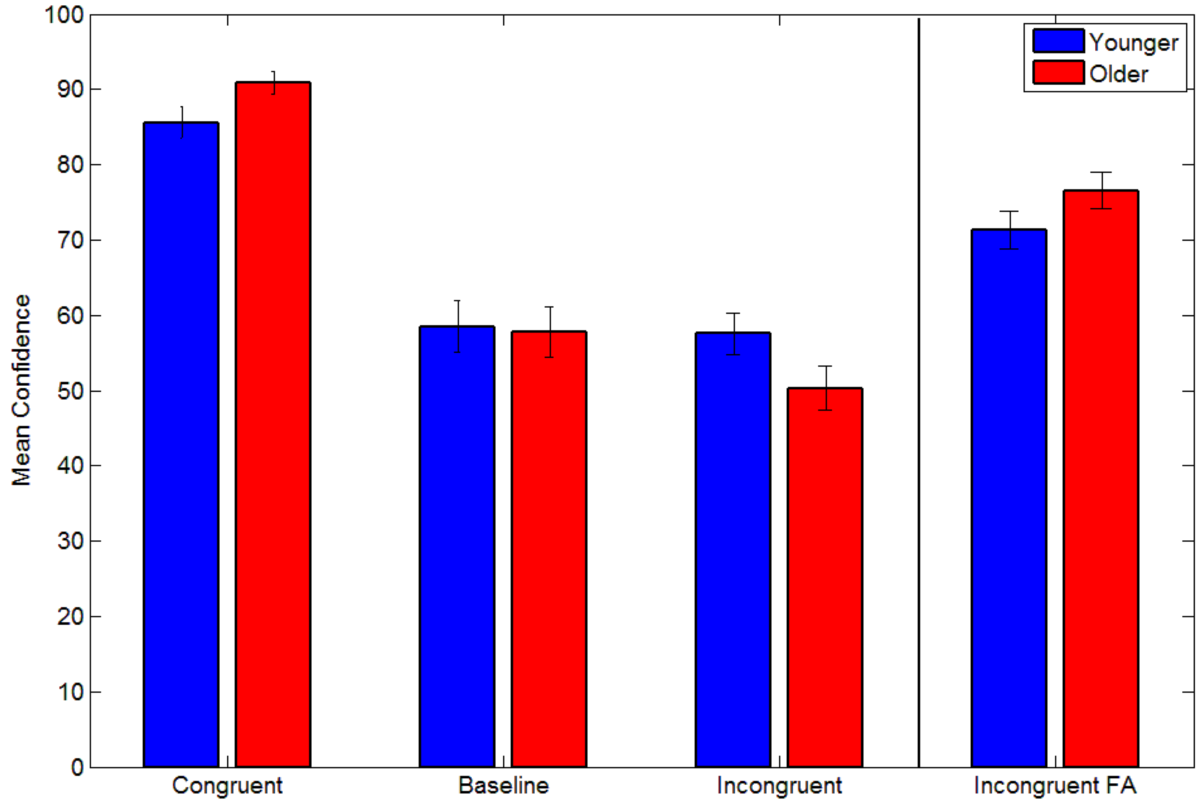

Figure 2. Mean confidence of correct identification's and incongruent false alarms made for each trial type. Mean confidence for correct identifications of congruent, baseline and incongruent conditions are plotted to the left of the dividing line. Mean confidence of incorrect responses, predicted by context, on incongruent trials (false alarms) are plotted to the right of the dividing line.

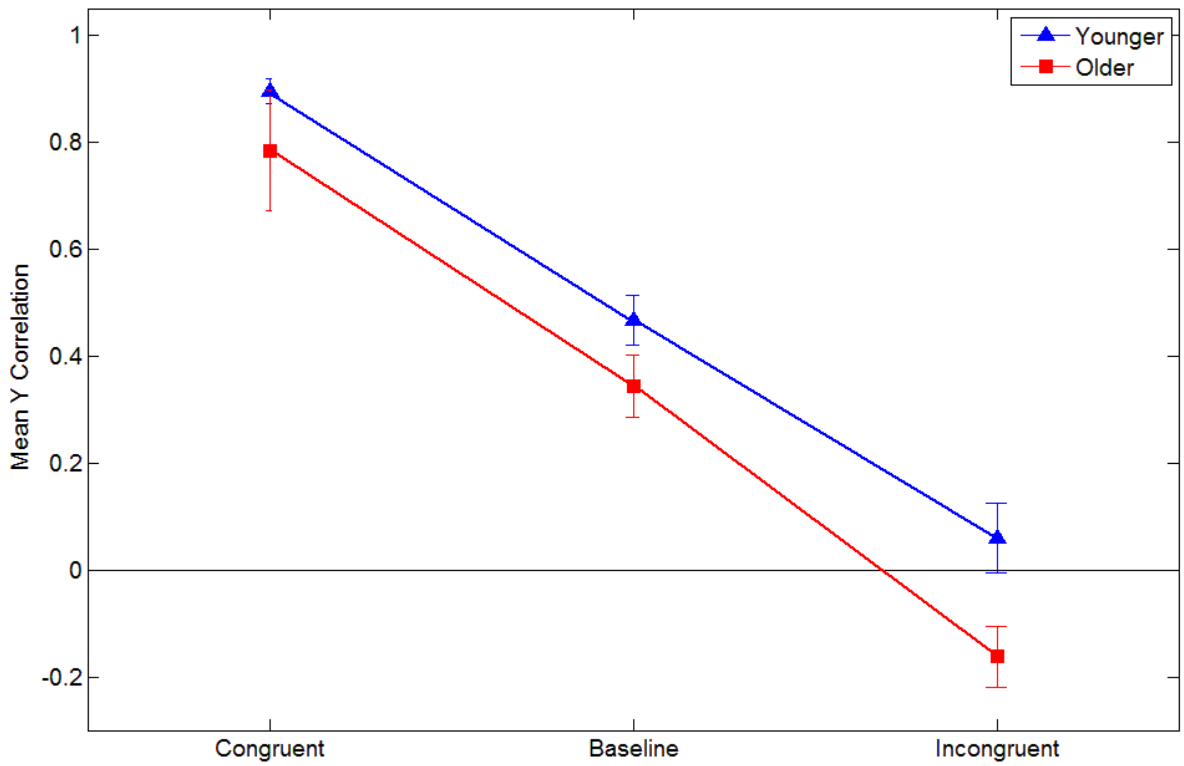

Figure 3. Gamma (γ) correlation data for each trial type. Values above the zero line correspond to a positive relationship between confidence and accuracy (good monitoring), whereas values below the zero line correspond to a negative relationship between confidence and accuracy (poor monitoring).

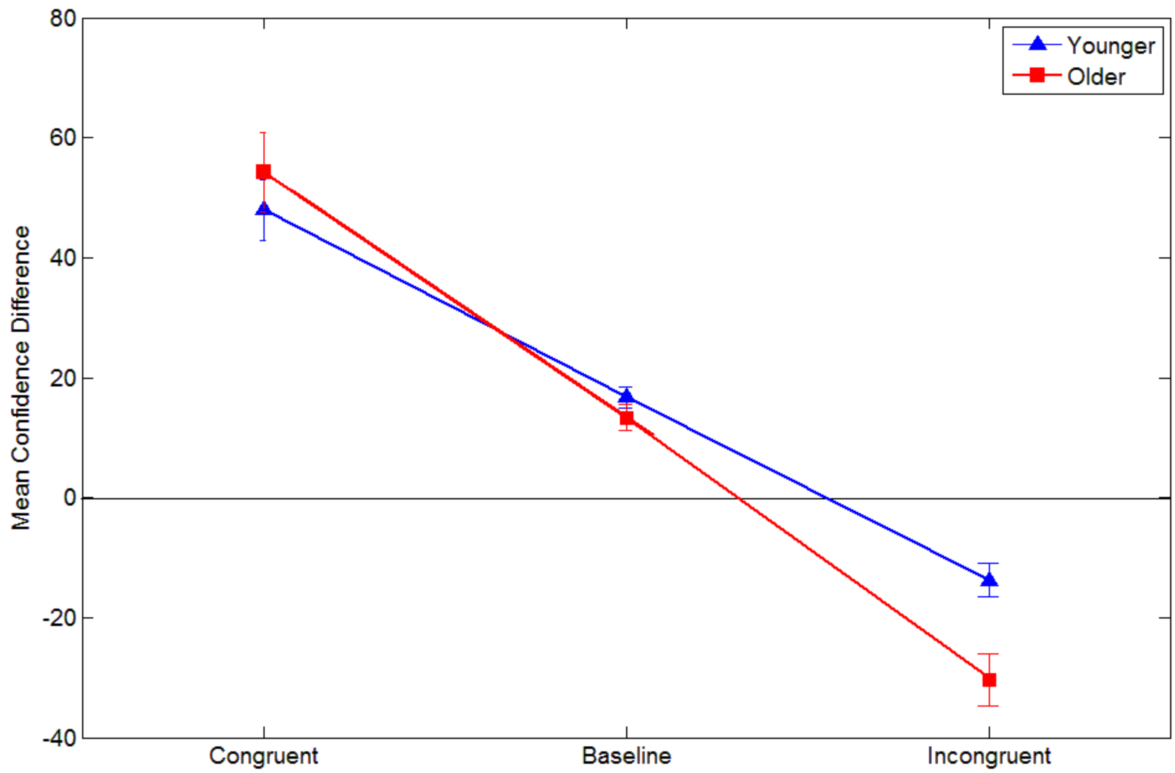

Figure 4. Mean confidence difference for each trial type. Values above the zero line correspond to a positive relationship between confidence and accuracy (good monitoring), whereas values below the zero line correspond to a negative relationship between confidence and accuracy (poor monitoring).

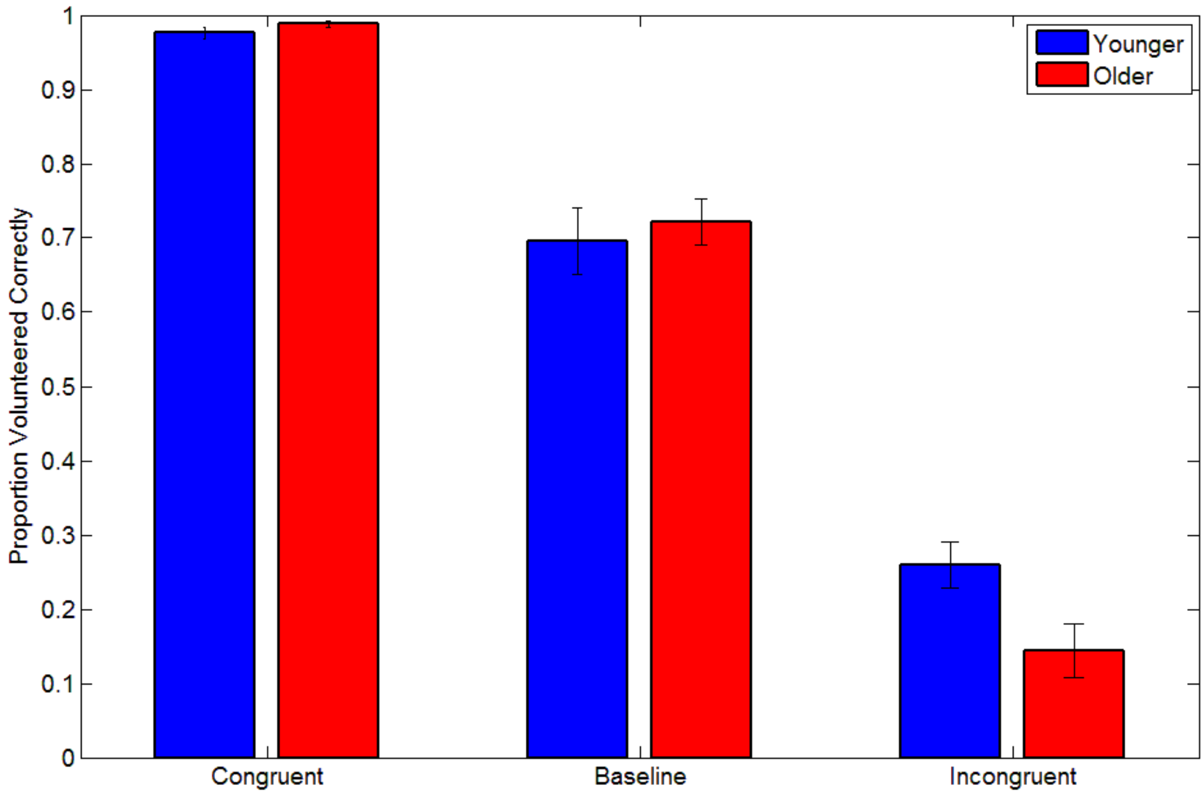

Figure 5. Proportion of correctly volunteered responses by trial type.