January 2010

Reducing False Hearing In The Elderly: Variability Training In Meta-Audition

Chad Rogers
Washington University in St. Louis

WASHINGTON UNIVERSITY
Department of Psychology
Dissertation Committee:

Larry Jacoby, Chair

Dave Balota

Henry Roediger, III

Mitchell Sommers

REDUCING FALSE HEARING IN THE ELDERLY: VARIABILITY TRAINING IN META-AUDITION

by

Chad Steven Rogers

A dissertation presented to the

Graduate School of Arts and Sciences

of Washington University in

partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

December 2010
Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Reducing False Hearing in the Elderly: A First Pass at Variability Training in Meta-audition.

by

Chad Steven Rogers

Doctor of Philosophy in Psychology

Washington University in St. Louis, 2010

Professor Larry Jacoby, Chairperson

A large body of work supports the conclusion that older adults derive more benefit than young adults from the addition of contextual information in speech perception (Hutchinson, 1989; Pichora-Fuller, Schneider, & Daneman, 1995; Pichora-Fuller, 2008). More recent work by Rogers, Jacoby, and Sommers (in prep) showed that when contextual information and sensory information favored competing responses, older adults were more likely to falsely “hear” the word favored by context. The current research describes two experiments that attempt to mitigate this age-related increase on contextual reliance. Experiment 1 assessed whether the effects of context in Rogers, Jacoby, & Sommers (in prep) were a result of repetition or semantic priming. The results of that experiment revealed that repetition of semantically-associated pairs did not increase false hearing, which supported the notion that context effects were a result of semantic priming. Experiment 2 described two variability-based training procedures aimed at reducing false hearing in older adults. The results showed that while variability-based training did not reduce false hearing to a greater extent than a practice-without-variability control group, age groups differed in their sensitivity to variability. Implications for further training, as well as practical implications for hearing aid users are discussed.

Acknowledgements

It is my pleasure to recognize the many people who have helped me through this process. This work would not have been possible without the generous guidance, wisdom, and encouragement of my advisor, Larry Jacoby and collaborator Mitchell Sommers. I would like to thank the other members of my committee: Dave Balota and Roddy Roediger for being inspiring teachers and for providing helpful comments and suggestions throughout my graduate career. This work was made possible by Washington University's Dean's Dissertation Fellowship, and Larry Jacoby's Departmental 9 Fund.

Thanks also to Carole Jacoby, Anna Dinndorf, Jennifer Schuitema, and Denise Rose for collecting and coding data and to Ben Anderson, Julia Feld, Nathan Rose, Dan Howard, and McKenzie Ballou for a wealth of technical and moral support. I am very grateful to my parents, Sharon and Mark Pennington and Charles and Cynthia Rogers, for nurturing and celebrating my intellectual growth and to Andrea Morris for keeping such good company while I find my way. I would also like to dedicate this thesis to my late grandmother, Dianne Elizabeth Rogers.

Table of Contents

List of Tables	vii
List of Figures	ix
Chapter 1: Introduction and Literature Review	1
Aging and the Facilitative Use of Context in Speech Perception	3
Aging and False Memory	7
False Hearing in Older Adults	10
Training Reductions in False Hearing	12
Overview of Experiments	18
Chapter 2: Methods for Experiment 1	20
Materials and Participants	20
Procedure	22
Chapter 3: Experiment 1 Results and Discussion	29
Baseline Trials	29
Congruent and Incongruent Trials	33
Summary	44
Chapter 4: Methods for Experiment 2	46
Materials and Participants	46
Procedure	49
Chapter 5: Experiment 2 Results and Discussion	53
Pretest and Training Analyses	54
Sensitivity to Variability in Context Validity (CV vs. NV)	67
Sensitivity to Variability in Signal Quality (SV vs. NV)	80

Summary 96

Chapter 6: General Discussion 98

Is Semantic Priming Responsible for False Hearing? 98

Implications for Training Reductions in False Hearing 100

Age Differences in Sensitivity to Variability 101

Conclusion 104

References 106

Appendix: List of Stimuli 115

List of Tables

Table 1. Results from baseline trials from Experiment 1. 31

Table 2. Hit and false alarm rates for congruent and incongruent trials in Experiment 1 35

Table 3. Mean confidence in congruent hits and incongruent false alarms in Experiment 1. 37

Table 4. Calibration bias for congruent and incongruent trials in Experiment 1. 39

Table 5. Monitoring data for incongruent trials in Experiment 1. 39

Table 6. Changes in proportion correct from forced to free report for congruent and incongruent trials in Experiment 1. 43

Table 7. Baseline hit rates at pretest and posttest in Experiment 2. 61

Table 8. Incongruent false alarm and volunteered incongruent false alarm rates during pretest and posttest in Experiment 2. 62

Table 9. Composite measures for sensory-based and context-based responding during pretest and posttest in Experiment 2. 65

Table 10. Hit and false alarm rates during the training phase for CV and NV participants in Experiment 2. 70

Table 11. Mean confidence in congruent hits, baseline hits, and incongruent false alarms during the training phase for CV and NV participants in Experiment 2. 74

Table 12. Calibration and monitoring data during the training phase for CV and NV participants in Experiment 2. 76

Table 13. Sensory-based and context-based responding composite for NV and CV participants during the training phase of Experiment 2. 79

Table 14. Hit and false alarm during the training phase for SV and NV participants in Experiment 2. 83

Table 15. Calibration and monitoring data during the training phase for CV and NV participants in Experiment 2. 91

Table 16. Sensory-based and context-based responding composite for NV and SV participants during the training phase of Experiment 2. 95

List of Figures

Figure 1. Schematic of conditions in Experiment 1. 28

Figure 2. Design schematic for Experiment 2. 50

Figure 3. Mean confidence ratings during the training phase for young NV and SV participants in Experiment 2. 86

Figure 4. Mean confidence ratings during the training phase for older NV and SV participants in Experiment 2. 87

Chapter 1: Introduction and Literature Review

During a hike in coastal Michigan, I approached a large fir tree to take a picture. A small songbird that was nesting in the tree started swooping 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 alright?!" After some discussion, it became clear that she was absolutely sure 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). My colleagues and I (Rogers, Jacoby, & Sommers, in prep) have demonstrated 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. We refer to the subjective experience of audition as meta-audition.

When perceiving a spoken word in context, listeners can base their perceptual experience on two distinct sources of information: sensation and context. 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, as well as 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. In the current example, sensory and contextual information are incongruent in that the sensory information strongly suggests "lark" whereas the contextual information strongly suggests "shark". Older adults are more likely to experience false hearing in these incongruent situations, even when young and older adults are equated on auditory sensitivity as measured by the ability to identify words heard in a neutral context (Rogers et al., in prep).

Why are older adults more likely to falsely "hear" in these situations? A simple explanation is that older adults are more reliant upon contextual information than young adults when comprehending speech (Rogers et al., in prep). This reliance may result from greater experience with listening in degraded listening conditions (Pichora-Fuller, 2008). This reliance may be very useful in situations where context is facilitative, but not in situations where context is incongruent with sensory information. In these incongruent situations, it is better to respond on the basis of sensory information, and not be influenced by the misleading context. Older adults' greater reliance on context could reflect an inability to constrain to sensory information. It could also reflect a resistance on the part of older adults to be flexible in terms of using both sensory and contextual information. One major goal of this dissertation is to provide a "first pass" at a training procedure to reduce false hearing. Rogers et al. (in prep) informed participants that context information could be misleading, but did not employ a training procedure to reduce false hearing in older adults. While most hearing training has been targeted toward improving sensory acuity in older adults (Sweetow & Palmer, 2005), the current training is aimed towards reducing false hearing.

First I will review prior work on context effects in speech perception that posed questions answerable by tests for false hearing, then I will review work in false memory that inspired the work in false hearing. After explaining how false hearing was observed in Rogers et al. (in prep), I will review work in the training literature relevant to a notion of training reduction in false hearing in older adults. Lastly, I will provide an overview of two experiments aimed at developing a training procedure for reducing false hearing in 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 may increase (e.g., Nittrouer & Boothroyd, 1990). In particular, older adults may rely more on top-down information such as semantic context to compensate for hearing impairment (Wingfield, Tun, & McCoy, 2005). When compared to young adults under degraded listening conditions (e.g., with moderate to high levels of background noise present), older adults' accuracy deficit diminishes significantly in the presence of meaningful semantic context (Dubno, Ahlstrom, & Horwitz, 2000; Hutchinson, 1989; Pichora-Fuller, Schneider, & Daneman, 1995; Sommers & Danielson, 1999).

One possible explanation for this age-related benefit from context use is that if older adults benefit more from the addition of contextual information than do young adults, then older adults possess a mechanism to utilize context to compensate for their age-related hearing loss (Pichora-Fuller, 2008). In an earlier paper, Pichora-Fuller, Schneider, and Daneman (1995) suggested that such a compensatory mechanism requires a reallocation of cognitive resources to process contextual information. In their study, age groups did not differ in their ability to identify words in noise that were presented with supporting contextual information, but older adults had greater costs than young adults on a concurrent working memory task. These costs were taken as evidence for older adults' greater processing of context. Sommers and Danielson (1999) argued such context use serves to reduce the demands on inhibition required for speech processing, which is deficient in older adults (Hasher & Zacks, 1988). They argued this reduced demand for inhibition increased lexical discrimination in older adults, which is the process of matching a word from an incoming speech signal with the correct representation of that word in the lexicon (Sommers, 1996).

An alternate explanation is that older adults are more reliant upon contextual information, rather than better able to discriminate in context. Reliance could be characterized as being more biased towards a particular response, regardless of its veridicality. Prior investigations of age differences in context effects have only explored situations in which the sensory signal and context are congruent (e.g., presenting "I was attacked by a shark") and thus favor the same response. To determine whether context influences reliance or lexical discrimination, one must include incongruent items for which context and sensory signal favor different responses. The benefit of including such incongruent items is that they allow for testing the possibility that context produces an increase in both congruent hits (correct responses) and incongruent false alarms (incorrect responses). Such a pattern favors the conclusion that older adults are more reliant upon context, as opposed to being better able to discriminate when context is available.

Being reliant upon contextual information is certainly adaptive if the context is not misleading, but flexibly switching focus to sensory information when context is likely to be misleading is also adaptive. Possibly, participants can focus their processing either on an *individual word* or on the larger unit of the *word-in-context*. That is, the individual word level and the word-in-context level serve as qualitatively different bases for auditory judgments, similar to the letter and word levels in visual perception as shown by the word superiority effect (e.g., Reicher, 1969; Wheeler, 1970). Focusing at the word level is a more effortful, inefficient 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. Given contextual cues, older adults might be less able or less willing to focus their attention on the word level, resulting in a greater rate of false hearing than young adults. Reliance upon the more global, word-in-context level facilitates performance when the word level and word-in-context levels are congruent (e.g., I was attacked by a shark) but can produce false hearing when the two levels are incongruent (e.g., I was attacked by a lark). Words perceived in natural environments are often accompanied by background noise, making attention at the context level even more likely because words are difficult to perceive. Attention at the more global, context level also explains the second aspect (apart from accuracy) of false hearing—high confidence in incorrect responses.

As stated earlier, consideration of the subjective experience of hearing has been largely neglected in favor of an emphasis on accuracy of responding. Among the exceptions is research on the phonemic-restoration effect. Classic research on phonemic restoration (Warren, 1970) found that when a word has a phoneme removed and replaced with white noise, participants do not report any removal of a phoneme and experience the word as being intact. In addition to the work on phonemic restoration, Jacoby, Allan, Collins and Larwill (1988) revealed an influence of memory on the subjective experience of hearing. In their study, participants first listened to a series of sentences and then listened to another series of sentences presented with background noise, including sentences from the first phase and new sentences. Participants were asked to judge the loudness of the background noise. Participants judged the noise to be less loud when it accompanied old sentences, even though the physical level of noise was identical for all sentences. A similar relationship between memory and audition was reported by Goldinger, Kleider, and Shelley (1999). They replicated Jacoby et al.'s (1988) finding of quieter noise judgments for old words and also demonstrated the reverse pattern; they found that new words presented in relatively soft background noise were more likely to be judged as "old" during a recognition memory test than new words presented in a louder background noise.

In both the listening and memory tasks, participants used multiple, qualitatively different bases for responding: prior presentation (memory) and the audibility of the word (background noise level). In terms of subjective experience, the contribution of memory influenced the subjective experience of background noise loudness even though audibility was unchanged. Additionally, a change in audibility influenced the subjective experience of *oldness* even though memory was unchanged. Jacoby and colleagues (e.g., Jacoby, Kelley, and Dywan, 1989) attributed effects of this sort to participants' reliance on a fluency heuristic (see also, Benjamin & Bjork, 1998; Kelley & Lindsay, 1993). Fluency of perception that is caused by differences in memory or context can be misattributed to differences in audibility, and vice versa. More generally, Rogers et al. (in prep) held that there are multiple levels, comprising different unit sizes, that contribute to perception and memory performance, and contributions from more global processing (e.g., word in context) can be misattributed to a contribution of more local processing (e.g., audibility of the word). I will return to this possibility when discussing the potential relationship between false hearing and false memory.

Aging and False Memory

Jacoby and Rhodes (2006) reviewed experiments demonstrating that older adults are much more prone to false memory than young adults and described those findings in terms of a dual-process model of memory. Many different experimental procedures have been used to elicit false memories in young and older adults (e.g., Loftus & Palmer, 1974; Lyle, Bloise, & Johnson, 2006; Roediger & McDermott, 1995; Thapar & Westerman, 2009). I focus on experiments done by Jacoby, Bishara, Hessels, & Toth (2005) because my false hearing experiments employ procedures similar to those they used to find age differences in false remembering. In their experiments, participants studied a list of associatively related word pairs (e.g., knee-bone) for a later test. Following the study phase, participants took a cued-recall test and judged whether they truly "remembered" each response that they reported as having been earlier studied or whether they were only "guessing." Test items for the cued-recall test consisted of a cue word and word fragment that could be completed with either the target (e.g., bone) or an alternate response (e.g., bend) that was also associatively related to the cue word (e.g., knee-b_n_). Just prior to each test item, a prime word was briefly presented that was either congruent (e.g., studied "knee-bone;" prime "bone"; test "knee b_n_") or incongruent (e.g., studied “knee-bone;” prime “bend”; test “knee b_n_”) with the studied pair. In all cases, however, the participants’ task was to respond with the studied pair (knee-bone).

Presenting a congruent prime increased both the probability of correct responding and claims of “remembering” for both young and older adults. More interestingly, presenting an incongruent prime word reduced memory accuracy more for older adults than for young adults, and older adults were much more likely to mistakenly report “remembering” studying the incongruent prime (i.e., false memory). In contrast, young adults were much more likely to indicate that they were “guessing” when they incorrectly claimed to have studied the incongruent prime. In one experiment, older adults were ten times more likely to falsely “remember” the incongruent prime than were young adults. These results were obtained even though older adults saw the pairs for a longer duration of time during the study phase than did young adults, in order to equate memory for the studied pairs.

Are older adults willing to act on their misled subjective experience by responding even when not forced to do so? Jacoby et al. (2005, Expt. 3) allowed participants to withhold non-confident responses (i.e. guesses) and thus reduce the probability of false recall (Koriat & Goldsmith, 1994; 1996). Older adults used their subjective experience as a guide for deciding whether or not to volunteer a response, as they were much less likely to take advantage of an opportunity to not respond than were young adults. This finding highlights the importance of subjective experience for older adults, as they were particularly willing to act on a high-confidence judgment even when it was made in error.

Hay and Jacoby (1999) used a procedure similar to that used by Jacoby, et al., (2005) to show that misleading associative context (in the form of proactive interference) can be a potent source of false remembering for older adults (see also Jacoby, Debner & Hay, 2001). To illustrate their findings they described the case of an elderly math professor who was unable to locate his return airline ticket after a conference. He purchased a new ticket and, upon arriving home, phoned his wife to pick him up at the airport. She responded that she would be unable to do so because he had driven their only car to the conference! The professor's action slip reflected a failure to recollect driving to the conference, instead relying on his usual habits that accompanied flying. In this vein, Hay and Jacoby described older adults' greater susceptibility to proactive inhibition using a dual-process model 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.

Jacoby and colleagues contend that the controlled processes necessary for supporting recollection change with age, rendering bias effects more influential and leading to older adults demonstrating a greater rate of false memories. Just as there are multiple bases for responding, there are also multiple bases for subjective experience in the form of confidence (Jacoby, Kelley & Dywan, 1989). The high accessibility of a response that results from more global factors such as associative context (e.g., Hay & Jacoby, 1999) can serve as a basis for the subjective experience of remembering and thus can result in false remembering in incongruent contexts. Similarly, a background noise can be judged as less loud when accompanied by a sentence that was previously heard (Jacoby, et. al., 1988). In both cases, subjective experience relies on a fluency heuristic that results in a change at one level (associative context or memory) being misattributed as having originated at another level (memory for a particular event or loudness of background noise).

As mentioned earlier, Rogers et al. (in prep) expected to obtain parallel results demonstrating that older adults are more prone to false hearing. As in false memory, they expected this greater susceptibility to result in older adults being less able or less willing to rely on an effortful, consciously-controlled basis for responding (sensory information). Instead, older adults should more heavily rely on a more automatic, global form of processing (contextual information). Relying on sensory information for "hearing" serves the same role as reliance on recollection does for "remembering", in that both serve as a valid basis for subjective experience. In contrast, reliance on associative context can serve as a source of false hearing just as it can serve as a source of false remembering. Similar to the case of the elderly math professor who left his car at the airport, older adults might rely on what is usually said rather than what is said on a particular occasion.

False Hearing in Older Adults

To study the effects of listening context for older adults in speech perception, Rogers et al. (in prep) conducted experiments in which context served as both a valid and invalid basis for responding. In that study, participants were trained on a series of paired-associates (e.g. BARN- HAY), with memory being assessed on a cued-recall test. After the recall test, participants were given a perceptual test in which a cue word was played in the clear, followed by a target word in noise.   Participants had to identify the second word. On congruent trials, the cue from training was paired with its old target (e.g. BARN-HAY). On incongruent trials, a cue from training was paired with an alternate word that was a phonological neighbor of its old target (e.g. BARN-PAY). The incongruent trials serve as a situation in which context and sensory information favor competing responses. The result was that older adults were much more likely than young adults to produce the response favored by context, both when it was a correct response (e.g. congruent trials) and when it was an incorrect response (e.g. incongruent trials).

Another novel feature of their procedure was to measure meta-audition by asking participants to rate their confidence on a 0-100% scale after every identification response, and choose which trials to volunteer for scoring. Such a procedure is used in memory tasks such as recall and recognition (Koriat & Goldsmith, 1996). Participants were urged to be confident and volunteer responses only on the basis of what they actually heard. This measure is important for false hearing, as a participant may produce a response favored by context, but do so with low confidence and choose not to volunteer the response. This is different from the case of false hearing where a participant falsely reports hearing the word favored by context, and then volunteers it.

Rogers et al. (in prep) found that that older adults were far more reliant upon context than young adults. When context was facilitative, older adults were more accurate than young adults. However, when context was misleading, older adults were more likely than young adults to incorrectly respond with the word favored by context. These differences in accuracy could not be explained in terms of guessing differences between age groups. In their procedure, guessing behavior should be accompanied by low-confidence responses. Rather, older adults were extremely confident when providing the response favored by context, even on incongruent trials. On incongruent trials, older adults reported being 100% confident on 39% of trials. These high confidence responses were nearly always the incorrect response favored by context, thus the relative accuracy of these responses was quite low (<5%). In contrast, young adults only reported being 100% confident on 13% of incongruent trials, and with a relative accuracy of 41%.

The false hearing work is of theoretical interest because it bridges the fields of memory and speech perception. The finding of greater false hearing in older adults mirrors the work in aging and false memory. Older adults are more likely than young adults to falsely remember an event when source-constrained retrieval or recollection is required as a basis for veridical responding (Norman & Schacter, 1997; Jacoby et al., 2005). Though these two experiments use very different methodologies, they both elicit false memories by creating a highly accessible but invalid global basis for responding (e.g. familiarity). In order to avoid the misleading influence of these global bases, participants must only respond on the basis of retrieval from the prior study episode. The analogy to false hearing is that global context is a misleading basis for responding, and that older adults falsely hear more because they respond on the basis of sensory information to a lesser extent than young adults.

Training Reductions in False Hearing

Greater false hearing in older adults could arise from an age-related deficit in the ability to constrain to a particular source of information, as has been suggested for false memory (Jacoby, et al., 2005). Another possibility is that age differences in false hearing may reflect a lack of flexibility in responding; older adults may be so accustomed to using context in naturalistic listening situations that they are unable or unwilling to switch towards using sensory information as a basis for responding. Age-related deficits in flexibility of reporting have been shown in memory (Aizpurua & Koutstaal, 2010; Koutstaal, 2006). Thus, training experiments targeting constraint and flexibility may be particularly informative for training and false hearing.

On the issue of training constraint to veridical sources of information, a set of studies targeting recollection training (Jennings & Jacoby, 2003; Jennings, Webster, Kleykamp, & Dagenbach, 2005) have demonstrated promising outcomes, including robust far transfer. In their procedure, older adults study a list of words for a later recognition memory test. During the test, the new foils are repeated at varying intervals (lags). Participants are instructed only to respond "yes" to words they read on the study list, and "no" to all other words. The repeated foils are familiar, but familiarity alone is not a valid basis for responding, and recollection of their presentation during the test list is required for them to be rejected. At first, older adults could only reject repeated foils (also called *avoiding repetitions)* at very short (1 or 2 item) lags, but after a training-to-criterion training procedure, were able to increase the lags to 28 intervening items when training was completed. At that 28-item intervening lag, no difference was found between young and older adults. Equally important, those that received recollection training also showed a significant pre- to post-training gain on four different transfer tasks, including a working memory (n-back) task, self-ordered pointing, source monitoring, and the Digit Symbol Substitution Test.

Similar transfer effects to those found in recollection training (Jennings & Jacoby, 2003; Jennings, Webster, Kleykamp, & Dagenbach, 2005) were obtained using procedures targeting cognitive flexibility (Tranter & Koutstaal, 2008). In the Tranter and Koutstaal (2008) procedure, older adults randomly assigned to an experimental treatment group practiced a varied set of novel and engaging tasks that encouraged flexible problem-solving. These tasks consisted of activities such as word-logic problems, construction tasks using household objects, and critiquing very unfamiliar types of music (e.g., Tuvan throat music). This experimental treatment group was compared to a control group that did not perform any of these activities, but came to the laboratory for social gatherings. Participants in the experimental treatment group got better with practice at their novel tasks, but also showed much better pretest to posttest gains on tasks measuring fluid intelligence such as Cattel's Culture Fair test (Cattell & Cattell, 1960) and the WAIS-R blocks (Wechsler, 1981). Tranter & Koutstaal (2008) concluded that their novel and engaging tasks encourage creative and flexible thinking, which are important for fluid intelligence; the type of intelligence that is most diminished in older adults (Horn & Cattel, 1967; Schaie, 1994). Tranter & Koutstaal (2008) and Jennings et. al. (2005) both argued that the reason why they found such broad transfer effects is that they trained processes that are used by many cognitive systems. Such increased flexibility could generalize to false hearing as well.

A different kind of test for age differences in flexibility is to test for age differences in responsiveness to different payoff matrices as measured by shifts in response criterion (Green & Swets, 1966). In a recognition memory task, response criterion refers to the bias to report an item as "old," which is distinguished from discriminability, which refers to the ability to distinguish between old and new items. Baron and Surdy (1990) tested for age differences in shifting criterion for responding in a memory task. They trained young and older adults on a continuous recognition procedure similar to that used by Jennings and Jacoby (2003), but shifted payoff matrices within-participants. The payoff matrices varied from favoring “old” responses (e.g., +9 for a hit, -1 for a false alarm), favoring “new” responses (e.g., +9 for a correct rejection, -1 for a miss), to neutral (+5 for hit/correct rejection, - 5 for false alarm/miss). Participants completed an initial set of 12 neutral blocks, then six blocks favoring “old” responses, six blocks favoring “new” responses, and then 12 final neutral blocks. Their results were that older and young adults’ discriminability benefited significantly from training. Also, both older and young adults changed their criterion in line with the shifted payoff matrices, but older adults shifted to a less extent than young adults. In fact, one older adult (participant O4) was completely insensitive to changes in the payoff matrix, indicative of some inability or unwillingness to be flexible in reporting strategy.

Training flexibility in the way Baron and Surdy (1990) did with payoff matrices is a very attractive target for false hearing training. Even though contextual information is usually helpful in natural listening environments, to be flexible enough to switch to sensory information when context is misleading is important for correct hearing. To encourage such flexibility, rather than shifting payoff matrices as done by Baron and Surdy (1990), the training experiment the current thesis (Experiment 2) included situations in which contextual information varied in terms of its validity (e.g. facilitative on 70% or 30% of trials) as well as situations in which sensory information varied in terms of signal quality (e.g., more or less noise masking relative to control). As in the Baron and Surdy (1990) procedure, participants completed neutral pretest and posttest blocks. Thus, two research questions were addressed. First, does exposure to variability encourage flexibility and/or reduce false hearing relative to a no variability control? Second, do older and young adults differ in their sensitivity to that variability, as was shown by Baron & Surdy (1990)? While Baron & Surdy (1990) were only concerned with accuracy measures, Experiment 2 includes measures of meta-audition as well.

So, is exposure to variability a viable training option for reducing false hearing? Kerr and Booth's (1978) classic beanbag experiment has shown that practice with variability improves accuracy in a motor task. In that study, 8-year olds and 12-year olds were trained to throw miniature beanbags at a target placed at a criterion distance. The participants were divided into two groups, the low variability group, and the high variability group. Before throwing at the target, participants were given four practice throws. For the low-variability group, the four practice throws were made at the criterion distance. For the high-variability group, the four throws were varied around the criterion distance, but were never made at that particular distance. At test, the high-variability participants were more accurate at the criterion distance, although they had never thrown at that distance. Schmidt and Bjork (1992) later proposed that the benefits of variable practice may extend well past those found in motor learning, and into general cognitive training (see also Kramer, Larish, & Strayer, 1995; Goode, Geraci, & Roediger, 2008). In hearing, Barcroft & Sommers (2005) found that those that those who practice with acoustic variability while learning foreign language vocabulary words show better retention and comprehension on a later test than those who practice with the criterion test.

Practice with variability may serve to reduce false hearing in older adults. If given practice in a situation where the noise levels vary from poor to good SNRs, participants may get a better sense of what the underlying bases for responding are (e.g., context and sensory information). In the Rogers et al. (in prep) procedure, participants were instructed to respond only on the basis of what was heard in the noise. In terms of confidence ratings, participants given practice with variability may commit fewer high confidence errors than a group only given practice at the criterion noise level a la Kerr and Booth (1978).

Variability training may serve as a better first pass at training reductions in false hearing than the intensive procedures such as those used by Baron and Surdy (1990), Jennings et. al (2005), and Tranter and Koutstaal (2008). To observe increased recollection in older adults, participants in the Jennings et. al. (2005) study were trained for six sessions across three weeks. The Tranter and Koustaal (2008) experimental training procedure consisted of 10-12 weeks of at-home practice sessions, complemented by three in-lab training sessions, in addition to pretest and posttest. Baron and Surdy (1990) trained participants during 2-50 minute training sessions per day for a total of 40 hours. These procedures require substantial costs that may not be necessary to reduce false hearing. Unlike recollection, increasing sensory information does not require training. In hearing tasks such as those used by Rogers et al. (in prep), enhancing sensory information is as simple as decreasing the amount of noise given to each participant. If given experience in a situation where sensory information is more available, then participants may learn to better constrain to avoid the potentially misleading nature of context. In the case of flexibility training, training in situations where participants practice flexibility across bases of responding (e.g., sensory information and context) may prove sufficient to reduce false hearing.

Overview of Experiments

Experiment 1 was designed to answer a critical question regarding the method used by Rogers et al. (in prep): is semantic priming enough to elicit false hearing, or does it require an experiential component? Recall that in the Rogers et al. (in prep) procedure, participants were trained on a series of cue-target pairs (e.g., BARN – HAY) before the perception test as a way to build context experientially. During that training, participants saw each cue-target pair five times before completing a cued-recall test. As all cue-target pairs were semantically related, it is not known whether the actual presentation of the items at training is necessary to elicit false hearing. This is important because such training does not represent a naturalistic listening situation, and because of the potential confound of age-related changes in associative memory (Naveh-Benjamin, 2000). In the first experiment of this dissertation, the number of presentation exposures of cue-target pairs was manipulated. Cue-target pairs appeared during training, zero, three or five times. False hearing was assessed in young and older adults as a function of presentation count. The lowest number of presentations needed to obtain a reliable effect of false hearing was used as the number of training exposure during the second experiment.

Experiment 2 was designed to test if a single-session training procedure incorporating practice with variability is better for reducing false hearing in the elderly than practice without variability. This experiment utilized a pre-post test design similar to that of Baron and Surdy (1990). Between pretest and posttest, young and older adults received training with feedback. There were three different training types: signal variability practice (SV), where participants were exposed to two different levels of noise, context variability practice (CV), where participants were exposed to situations in which context was predominantly facilitative or predominantly interfering, and no variability practice (NV), where noise level and context were the same as in the pretest and posttest. Note the similarities between this design and that described by Kerr and Booth (1978). Participants in the no variability practice group received more practice on the criterion test, but it was predicted that they would be more likely to falsely “hear” at posttest than either variability training group. It was also predicted that participants in the variability practice groups would be more likely to respond on the basis of sensory information than those in the no variability group, and thus show less confidence in responses favored by context, decreases in congruent hit and incongruent false alarms rates, and be less likely to volunteer incongruent false alarms for scoring at free report.

In addition to assessing the efficacy of the training procedures, Experiment 2 was designed to test if older adults differed from young adults with respect to their sensitivity to the variations present in the signal and context variability groups. If older adults were more likely to base responding on contextual information, would their confidence judgments be as sensitive as those of young adults when there was variability in the level of noise? Rogers et al. (in prep) found that older adults changed their level of confidence under heavier noise masking in a manner comparable to young adults, but that was done without adjusting the level of noise for each participant. When set to a participant’s individual 50% threshold, and thereby removing the contribution of age difference in hearing ability, a different pattern of age differences may emerge. What about when the context varied in terms of its validity? This question bears upon whether older adults are able to take note of situations where context is often misleading.

Chapter 2: Methods for Experiment 1

Materials and Participants

Participants. Eighteen undergraduate students were recruited through the Washington University participant pool in St. Louis, MO. They received either $10 or course credit for their participation. These younger participants ranged in age from 18-21 years old (*M*=19.28, *SD*= 1.26). Eighteen older adults were recruited through the Washington University Older Adult subject pool. These older participants ranged in age from 65-83 years old (*M*=70.83, *SD*= 4.69). The older participants were volunteers from the St. Louis community and also received $10 for their participation. Vocabulary scores from the Shipley Institute of Living Scale (Shipley, 1946) were obtained from each participant at the end of the session to assess vocabulary proficiency. Older adults had better performance on the Shipley scale (Young *M*= 33.28, Older *M*= 36.06), $t(34) = 3.29$, $p<.01$.

Materials. Ninety-four 3-word sets (cue, target, alternate) were created to serve as the test items for congruent and incongruent trials (e.g. context items). In these 3-word sets, the cue word and the target were semantically related. An additional 32 2-word sets (cue, target) were created to serve as the test items for baseline trials, where the cue word was not semantically related to the target. For all sets, the targets were English language consonant-vowel-consonant words (CVCs, with respect to phonology, not orthography) with corresponding alternates (also CVCs) with one consonant (either initial or final) changed along place of articulation. For example, the cue word "row" is semantically related to target CVC "boat," but not alternate CVC "goat," which differs phonologically from "boat" by one phoneme that shares the same voicing (e.g. voiced) and manner of articulation (e.g. plosive), but differs in place of articulation (e.g. velar). This was a context item. An example of a baseline 2-word set is “brief” which served as a cue word that is not semantically related to target word “lit.”

These materials were generated by using a list of all CVCs in the English language and University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) to find all CVCs that: a) had a word with a forward association strength to the target of .2 or greater that could be used as a cue; and b) had a potential CVC alternate that was an English word. Once all viable target CVCs were identified, the most phonologically confusable CVC (determined by data from the Washington University Neighborhood Activation website) that satisfied the aforementioned criteria was chosen as the alternate CVC. The 94 3-word sets with the highest forward association strength from cue to target were chosen to serve as materials for the congruent and incongruent trials. Those 94 3-word sets were divided into six component sublists of 15 which were balanced for cue, target, and alternate word frequency, phonological confusability between target and alternate, as well as the average forward and backward association strengths (FSG M=.55, BSG M=.25) between the cue and the target. The remaining four sets were used for practice trials. The component sublists were rotated through conditions across participants, such that no items were repeated within participants, but occurred equally in each of the appropriate conditions (e.g., congruent 0x, congruent 3x, congruent 5x, incongruent 0x, incongruent 3x, incongruent 5x).

Thirty-two additional CVCs were chosen to serve as baseline targets. Baseline cue words were chosen to match the lexical and phonological characteristics of the cue words used for congruent and incongruent trials. Baseline cues had no apparent semantic relationship to the target word. Thirty of these words were used during the actual perceptual test phase, the other two were used for practice.

The auditory stimuli were words from the above sets and were recorded with a Shure microphone in a double-walled sound-attenuating booth. The words were spoken by a male speaker with no heavy accent, and rms amplitude of all stimuli was equated to 65 dB SPL. Stimuli were masked by noise by taking the clear speech file (65dB SPL) and mixing it with a corresponding 6-talker babble noise file using E-Prime (PST, 2003) software. Mixing and leveling were done on a single PC-compatible computer with a Sound Blaster sound card. The signal to noise ratio varied for each participant and corresponded to the 50% speech reception threshold (SRT, ASHA, 1988) obtained during the titration procedure.

Procedure

Titration. Upon arrival, informed consent was obtained from all participants, and the experimenter administered a background questionnaire. A modification of the American Speech-Language-Hearing Association's recommended procedure (ASHA, 1988) was used for obtaining a SRT to establish the signal-to-noise ratio (SNR) used during the perceptual test phase for each participant. The ASHA procedure was used to measure an individual's threshold for speech instead of pure-tone audiometry.

The titration procedure used in this dissertation differed from the recommended ASHA procedure in a number of key ways. First, the task used for reporting the word in noise in the ASHA procedure was a recognition judgment from a pre-familiarized set of words. In Experiment 1, an open set cued identification task (e.g., "Say the word you heard out loud.") was used. Second, the ASHA procedure recommended the use of spondaic words as the test material for obtaining a SRT. Spondaic (also known as "spondee") words are two syllable words with equal stress on both syllables. Spondee words also are compound words where each syllable is a word as well, such as "baseball" or "hotdog." However, the words masked by noise in Experiment 1 were all monosyllabic, so monosyllabic words were used during the titration procedure. Third, ASHA procedures recommended the use of an audiometer and an audiometric booth for obtaining a SRT. For our titration procedure, testing was conducted under the same conditions as the rest of the experiment viz., in a quiet room with a calibrated set of Sennheiser headphones and a Sound Blaster sound card.

The exact procedure for obtaining a SRT is reported elsewhere (ASHA, 1988), but in brief, the procedure can be divided into two phases; the preliminary phase and the test phase. The goal of the preliminary phase was to establish a starting level with broad (10-dB changes in SNR) steps before honing down to smaller steps (2 dB changes in SNR) to obtain the threshold. To begin the preliminary phase, the experimenter presented a word at an SNR of 35 to the participant. If the participant identified the word correctly, the experimenter reduced level of the noise 10 dB (i.e., a -10 change in SNR) until the participant gave an incorrect response. If the participant did not identify the word correctly, the experimenter increased the level by 20 SNR and started again. When one word was missed, a second word was presented at the same level. The process of descending in 10 SNR steps was completed once the participant missed two words at the same level. Once that occurred, the level was increased by 10 SNR and that was the starting level obtained.

For the test phase of the titration procedure, two words were presented at the starting level and descended in 2 SNR steps, playing two words at each step. The procedure continued as long as the participant identified four out of the first five words. If the participant failed to do this, the starting level was increased by 10 SNR and the test phase restarted. The 2 SNR descent procedure was completed once the participant missed five of the last six words presented. Once completed, the threshold was calculated with the formula:

$$SRT = \text{starting level} - \#\ \text{correct} + 1.$$

Familiarization. During the familiarization phase, participants were seated in front of a PC and read instructions that they would study a list of word pairs. Their task was to remember the words for a later memory test, and given the first member of a pair, they should be able respond with the second member of the pair. To facilitate performance, participants were encouraged to think about how the cue and target were related to one another.

During the familiarization phase, participants were presented with a subset of 90 cues and targets from the 3-word sets used for the context trials mentioned in the Materials section. For each pair, the cue word (e.g. ROW) was presented on screen and then 100 ms later, the same word was presented aurally through the headphones. Fifty ms after the offset of the spoken cue word, the target word (e.g. BOAT) was presented visually on screen and then 100 ms later, the same word was also presented aurally through the headphones. After the offset of the spoken target word, a single asterisk "*" was presented on the screen for 500 ms before the next pair was presented. Only four of the context trial sublists of 15 3-word sets were presented during the familiarization phase: those assigned to the congruent 3x, congruent 5x, incongruent 3x, and incongruent 5x conditions. Those belonging to the congruent 0x and incongruent 0x conditions were not presented during the familiarization phase. During familiarization, only cues and targets were presented, never the alternates. Baseline sets were also never presented during familiarization. During the familiarization phase, all pairs were presented in a pre-randomized order to ensure a consistent ratio of 3x and 5x items throughout the familiarization phase.

The final component of the familiarization phase was a 60-item cued recall test to assess training. On each trial, the cue word was presented visually and aurally in a manner congruent with the training, but with a question mark following the cue word (e.g. ROW -?). Participants had five seconds to provide the target word, and were encouraged to guess if they did not know. After the participants provided a response or timed out, the target word was presented visually adjacent to the cue word (e.g. ROW – BOAT), and the word was played through the headphones. All participants were at 80% recall or above.

Open Set Cued Identification Task. After the familiarization and its accompanying cued-recall test, participants completed 120 trials of an open-set cued identification task. During this task, participants were instructed that they were again going to be hearing a series of cue-target pairs, but that the target would be masked by noise. After the word in noise was played, a question mark, "?" appeared on the screen prompting the participant to say the word that was presented in the noise aloud. Participants were warned that some of the pairs in the perception test phase were the same as the pairs in the familiarization phase (e.g. ROW – BOAT), but also that some of the pairs were different (e.g. ROW – GOAT). Participants were explicitly told to respond only on the basis of what they heard, and not on the basis of what they had earlier learned, or on the basis of semantic association between words. This last point was printed in CAPS on the computer screen and was emphasized by the experimenter when recapitulating the instructions.

After word identification, participants were instructed to provide a numerical response that corresponded to how confident they were that they provided the correct response. The scale for this judgment ranged from 0-100, where 0 corresponded to the lowest level of confidence and 100 corresponded to the highest level of confidence. After making a confidence judgment, participants were asked if they wanted to "volunteer" or "withhold" that trial. Participants were told that the computer was keeping score for their responses, but only for the trials volunteered. If they volunteered a trial, and identified the word in noise correctly, they received one point (+1). If they volunteered a trial, but did not identify the word correctly, they lost one point (-1). A withheld trial was a functional "pass" and regardless of whether or not the participant was correct, no change was made to their score. Participants were not shown the correct word after responding. They also did not see their score or a change in score after responding. Participants only got to see their score when the experiment was completed.

During the identification task, there were seven different types of trials: congruent 0x, congruent 3x, congruent 5x, incongruent 0x, incongruent 3x, incongruent 5x, and baseline trials. The 0x, 3x, and 5x manipulation refers the number of times a cue-target pair was presented during the familiarization phase. There were 15 trials of each type, except for baseline trials, of which there were 30. On congruent trials, the cue word was presented in the clear (e.g., ROW), followed by the accompanying target (e.g. BOAT) in noise. Incongruent trials began with a cue word from the familiarization phase, then the alternate (e.g. GOAT) was presented in the noise. Baseline trials had the cue word presented in the clear (e.g. BRIEF), followed by a word semantically unrelated to the cue masked by noise (e.g. LIT). Order of conditions in the perceptual test phase was pre-randomized, with the limitation that no more than three trials of a given type were presented in a row. For extra clarification, an illustration is provided in Figure 1.

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 1000 ms inter-stimulus 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 what word was being played over the headphones, but were offset so that they did not distract the participants as the word was being played. After participants opted to volunteer or withhold, the word "Ready?" appeared on the screen to which the participants had to respond "Yes" verbally before moving on to the next trial. Participants were given six practice trials before the beginning of the actual test.

Figure 1. Schematic of conditions in Experiment 1.

	Familiarization	Perceptual Test Phase	
	…	Baseline	BRIEF - *LIT*
0x	…	Congruent 0x	ROW - *BOAT*
		Incongruent 0x	ROW - *GOAT*
3x	ROW - BOAT ROW - BOAT ROW - BOAT	Congruent 3x Incongruent 3x	ROW - *BOAT* ROW - *GOAT*
5x	ROW - BOAT ROW - BOAT ROW - BOAT ROW - BOAT ROW - BOAT	Congruent 5x Incongruent 5x	ROW – *BOAT* ROW - *GOAT*

Note: Words presented within asterisks (*) were presented in noise.

Chapter 3: Experiment 1 Results and Discussion

Experiment 1 sought to test if repetition of a cue-target pair at training would increase false hearing. Rogers et al. (in prep) found strong evidence for false hearing for young and older adults when cue-target pairs were repeated five times during a familiarization phase. However, it is wholly plausible that the forward-acting associative strength of the cue to the target is sufficient to create the context effect observed in the congruent and incongruent trials of Rogers et al. (in prep) with such familiarization. In Experiment 1, some cue-target pairs were never presented during the familiarization phase (e.g., 0x trials), while others were presented either three times or five tiems). If evidence for strong false hearing was obtained on 0x trials, it suggests that the effect of context in this paradigm should be attributed to semantic priming, as the cue and target were always semantically associated. Unless otherwise specified, only effects found to be significant at the $\alpha<.05$ significance level that were not involved in a higher-order interaction are reported.

Baseline Trials

Before examining context effect differences in young and older adults, it is necessary to ensure that both age groups were presented age-appropriate noise masking at test. This manipulation check is helpful, because it allows for an examination of context effects in speech perception after accounting for age differences in hearing ability. To examine how older and young adults hear out of context, performance of both age groups on baseline trials were examined separately. Performance was assessed by examining the proportion of correct responses (e.g., hit rate), mean confidence rating ascribed to a hit, calibration bias, monitoring, and the probability of a hit subtracted from probability of a hit given it was volunteered (also called changes in accuracy from forced to free report; Sahakyan & Kelley, 2003). These measures from the baseline condition of Experiment 1 are presented in Table 1.

Baseline hit rate. As shown in Table 1, older and young adults did not differ in terms of their hit rates on baseline trials (Young *M* = .49, *SD* = .10, Older *M* = .48, *SD* = .07), $t(34)<1$, *ns*. Across age groups, hit rates did not differ significantly from .50, $t(35) = 1.00$, $p<.33$. This shows that the titration procedure was successful in controlling for age-related differences in hearing ability and securing a 50% level of performance in this open-set cued identification task.

Confidence in baseline hits. In terms of confidence, older adults trended toward being less confident in their baseline hits than young adults, (Young *M* = 55.22, *SD* = 19.10, Older *M* = 43.95, *SD* = 23.29), but the difference was not significant, $t(34) = 1.59$, $p<.13$. This was a trend found in Rogers et al. (in prep), where older adults were less confident in their hits than young adults in the baseline condition.

Calibration. Calibration bias is a measure of over/underconfidence. It is the mean difference between a participant’s average confidence rating and his or her accuracy. Positive values indicate over confidence, negative values indicate underconfidence. Values closer to zero indicate better calibration. On baseline trials, older adults trended towards being more underconfident on baseline trials than young adults, (Young *M* = -1.29, *SD* = 14.08, Older *M* = -11.82, *SD* = 17.96), $t(34) = 1.96$, $p<.06$. This trend is not surprising given equal performance across age groups with a trend towards older adults being less confident than young adults.

Table 1. *Results from baseline trials from Experiment 1.*

	Young		Older	
Baseline	*M*	*SE*	*M*	*SE*
Hit Rate	0.49	0.02	0.48	0.02
Confidence in Hits*	55	5	44	6
Calibration Bias	-1	3	-12	4
Monitoring	15	2	16	3
p(Hit|Vol)-p(Hit)	0.17	0.05	0.12	0.03

* Age difference was not significant, $t(34) = 1.59$, $p<.13$.

Monitoring. Participants' monitoring was assessed using confidence discriminability, which measures how participants' confidence differs when giving correct or incorrect responses. It is calculated by subtracting the mean confidence in an incorrect response from the mean confidence in a correct response. Positive values indicate good monitoring, as confidence in correct responses is greater than in incorrect responses. Negative values indicate poor monitoring. Young and older adults did not differ in monitoring on baseline trials, (Young M = 14.55, SD =8.22, Older M = 15.70, SD = 10.81), $t(34)<1$, *ns.* This finding also replicated Rogers et al. (in prep) where older and young adults had equivalent monitoring when sensory information was the only available basis for responding.

Changes in proportion correct from forced to free report. In Experiment 1, participants were given the opportunity to strategically regulate their proportion of correct responses by volunteering a subset of trials in which they felt confident enough to count toward a final score. The proportion of correct responses out of that subset of trials is called free report accuracy (Koriat & Goldsmith, 1996), which differs from the proportion of correct trials out of all trials completed, called forced report accuracy. By subtracting forced report accuracy from free report accuracy, one can measure the extent to which participants changed their accuracy from forced to free report (Sahakyan & Kelley, 2003). This is an important measure because it shows how well participants *act* on the basis of their subjective experience. The results from baseline trials for this measure are presented in Table 1, next to the descriptor "p(Hit|Vol)-p(Hit)." On baseline trials, young and older adults had similar gains from forced to free report (Young M = .17, SD = .11, Older M = .12, SD = .18), $t<1$, *ns*. One older adult was excluded from analysis for not volunteering any responses on baseline trials. This also replicated Rogers et al. (in prep, Experiment 2), where older and young adults did not differ in gains from forced to free report when sensory information was the only available basis for responding.

Aside from a tendency for older adults to be marginally underconfident on baseline trials, little evidence for age differences on baseline trials was found. Thus, it seems the titration procedure was successful at creating age-appropriate listening situations in the perception task. When titrated, older and young adults did not differ with respect to accuracy, monitoring, or gains from forced to free report. Now I turn to analyses that indicate the effect of repetitions on false hearing.

Congruent and Incongruent Trials

Congruent and incongruent trials were the trials in which context was available as a basis for responding. For congruent trials, context was facilitative, whereas for incongruent trials context was interfering. In Experiment 1, both trials occurred with the same frequency. For both congruent and incongruent trials, semantically associated pairs were presented 0, 3, or 5 times during the training phase.

Hit and false alarm rates. Based on prior work (Rogers et al., in prep), I hypothesized that older adults would have greater hits on congruent trials, but more false hearing on incongruent trials. Incongruent false alarms were a response favored by context. Note that in this study the rate of incongruent false alarms does not equal 1 minus the incongruent hit rate. In this open-set cued identification task, participants often reported words that were neither correct nor the word favored by context in the incongruent condition. Reporting in these analyses is limited only to incongruent false alarms as only they are indicative of context-based responding. The hit and false alarm rates for congruent and incongruent trials are presented in Table 2. The table shows that while older adults had greater hits than young adults on congruent trials (Young *M* = .84, Older *M* = .89), 1-tailed $t(34) = 1.86$, $p<.05$, older adults had fewer hits than young adults on incongruent trials (Young *M* = .46, Older *M* = .34), $t(34) = 2.47$, $p<.05$. The *t*-tests were qualified by a 2 (Age group: Young, Older) X 2 (Trial type: Congruent, Incongruent) x 3 (Presentation Count: 0x, 3x, 5x) mixed-model repeated measures ANOVA on hit rates that revealed a significant trial type x age interaction, $F(1, 34) = 9.37$, $MSE = .39$, $p<.01$, $\eta_p^2 =.22$. Presentation count had no significant effect or interaction with hit rates.

A comparison of age groups in Table 2 shows that older adults were much more likely than young adults to commit incongruent false alarms, (Young *M*= .31, Older *M*=.50) as indicated by a 2 (Age group: Older, Young) X 3 (Presentation Count: 0x, 3x, 5x) mixed-model repeated-measures ANOVA on incongruent false alarm rates with a significant main effect of age, $F(1, 34) = 11.63$, $MSE = .99$, $p<.001$, $\eta_p^2 =.26$. As with the congruent and incongruent hit rates, Table 2 shows little change in incongruent false alarms as a function of presentation count. Indeed, the ANOVA revealed no main effect or interaction with presentation count. Thus, repeated presentations during training did not increase the likelihood of reporting the response favored by context on incongruent trials. These data alone suggest that repeated presentations during the training phase were not necessary for inducing false hearing.

Table 2. *Hit and false alarm rates for congruent and incongruent trials in Experiment 1.*

	Presentation Count							
	0x		3x		5x		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent Hits								
Young	.82	.03	.85	.03	.85	.03	.84	.03
Older	.90	.03	.91	.03	.87	.03	.89	.03
Incongruent Hits								
Young	.47	.04	.44	.04	.47	.04	.46	.04
Older	.36	.04	.31	.04	.37	.04	.34	.04
Incongruent FAs								
Young	.33	.04	.28	.05	.32	.05	.31	.04
Older	.48	.04	.51	.06	.51	.06	.50	.05

Confidence in hits and false alarms. To assess the influence of repetitions on confidence, I restrict comparisons only to responses favored by context (e.g., congruent hits and incongruent false alarms). Confidence in these responses should indicate the extent to which participants based their confidence on contextual information when it was correct (e.g., congruent hit) and when it was incorrect (e.g., incongruent false alarm). Table 3 shows that older adults were more confident than young adults in both their congruent hits and incongruent false alarms, as supported by a 2 (Age group: Older, Young) X 2 (Trial type: Congruent, Incongruent) X 3 (Presentation Count: 0x, 3x, 5x) mixed-model repeated measures that revealed a significant main effect of age, $F(1, 33) = 6.59$, $MSE = 5197.53$, $p<.05$, $\eta_p^2 =.17$. Age did not interact with any variable. Across age groups, participants were most confident in their congruent hits at 5x presentation count, but for incongruent false alarms were most confident at 0x presentation count, which resulted in a significant trial type x presentation count interaction, $F(2, 66) = 4.23$, $MSE = 215.23$, $p<.05$, $\eta_p^2 =.11$.

While increased confidence as a function of repeated presentations for congruent hits is entirely consistent with the notion that repeated presentations build context effects on subjective experience, the reverse effect on incongruent false alarms is not. Instead, this finding suggests that repeated presentation of the response favored by context during the training phase may weaken the effect of context. Potentially, repeated presentations may have incremented participants' sensory requirement for responding with high confidence during the perceptual test phase. Thus, repeated presentations may actually reduce the likelihood of false hearing, relative to a situation where the response favored by context is never heard.

Table 3. *Mean confidence in congruent hits and incongruent false alarms in Experiment 1.*

	Presentation Count							
	0x		3x		5x		Overall	
Mean Confidence in	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent Hits								
Young	79	3	81	3	82	3	81	3
Older	89	3	92	2	93	3	91	3
Incongruent FAs								
Young	75	3	75	2	70	4	73	3
Older	84	3	83	4	82	4	83	4

Calibration. As was reported in the analyses of baseline trials, older adults had marginally less confidence in their hits than young adults when responding on the basis of sensory information alone. In contrast, Table 4 shows that for calibration bias older adults were better calibrated than young adults on congruent trials (Young *M* = -7, Older *M* = -3), but much more overconfident on incongruent trials, as revealed by a significant age x trial type interaction, $F(1, 34) = 6.91$, $MSE = 2924.08$, $p<.05$, $\eta_p^2 =.17$. This replicated Rogers et al. (in prep) and was consistent with the notion that older adults were more likely to be confident on the basis of contextual information than were young adults. This led to an advantage for older adults on congruent trials, but at a large cost on incongruent trials. Calibration bias also changed significantly as a function of repetition, as indicated by a significant main effect of repetition, $F(2, 68) = 3.59$, $MSE = 419.35$, $p<.05$, $\eta_p^2 =.10$. In contrast to the confidence data, there was not a significant interaction of trial type and presentation count. Post-hoc examination of the main effect of repetition revealed that 3x trials (*M*=14.65) were more likely to be associated with overconfidence than 0x (*M*=9.83) trials, no other differences were significant.

Table 4. *Calibration bias for congruent and incongruent trials in Experiment 1.*

	Presentation Count							
	0x		3x		5x		Overall	
Calibration	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent								
Young	-7	3	-7	3	-7	3	-7	3
Older	-5	3	-2	3	-1	3	-3	3
Incongruent								
Young	18	5	23	5	20	5	20	5
Older	34	5	45	5	38	5	39	5

Table 5. *Monitoring data for incongruent trials in Experiment 1.*

	Presentation Count							
	0x		3x		5x		Overall	
Monitoring	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Incongruent								
Young	-6	4	-3	5	3	5	-2	5
Older	-25	4	-11	5	-16	5	-17	5

Monitoring. Rogers et al. (in prep) found that when context was available as a basis for responding, older adults showed better monitoring than young adults when context was valid, but poorer monitoring than young adults when context was invalid. However, for Experiment 1 reporting of monitoring will be limited to only incongruent trials. Monitoring, as measured by confidence discriminability, cannot be calculated when accuracy is at ceiling or at floor. Twelve out of 18 older adults were at ceiling accuracy on congruent trials, and so confidence discriminability could not be reliably assessed. When analysis was restricted to incongruent trials, mean confidence in an incongruent hit was subtracted from the mean confidence when providing the response favored by context (e.g., an incongruent false alarm). Two older adults did not have any incongruent hits in the incongruent 3x condition, and were removed from analysis. One young adult was removed from analysis for not producing any incongruent false alarms in the incongruent 0x condition. The results of the analysis are presented in Table 5, which shows that that older adults had substantially poorer monitoring than young adults on incongruent trials, as revealed by a significant main effect of age, $F(1,31)=9.21$, $MSE=5947.77$, $p<.01$, $\eta_p^2=.23$. Importantly, reading Table 5 from left to right shows that across age groups, there was a significant main effect of presentation count, $F(2, 62) = 3.73$, $MSE = 799.35$, $p<.05$, $\eta_p^2=.11$, where monitoring was poorer on 0x trials (M=-15.28) than on 3x (M=-7.12) or 5x (M=-6.42). There was no significant interaction of presentation count with age, $F(2,62)=1.43$, $MSE=305.64$, $p<.25$, $\eta_p^2=.04$. These results suggest that 0x trials were the most effective for eliciting false hearing on incongruent trials across age groups.

Changes in proportion correct from forced to free report. Lastly I analyze how well participants were able to strategically regulate their accuracy in Experiment 1. Table 6 shows that the pattern of gains in proportion correct from forced to free report were quite different on baseline than on congruent and incongruent trials, where context could have been used as a basis for responding. Collapsing across congruent and incongruent trials, older adults showed poorer gains in proportion correct than young adults (Young *M* = .07, Older *M* = .01), as revealed by a significant main effect of age, $F(1,34)=9.07$, $MSE=.17$, $p<.01$, $\eta_p^2=.21$. However, the age group differences were smaller on congruent trials (Young *M* = .07, Older *M* = .05) than on incongruent trials (Young *M* = .06, Older *M* = -.03), which resulted in a significant age x trial type interaction, $F(1,34)=7.53$, $MSE=.06$, $p<.01$, $\eta_p^2=.18$.

With regard to repeated presentations, participants collapsed across age groups were similarly able to increase their proportion correct from forced to free report on congruent trials (0x *M*=.07, 3x *M*=.05, and 5x *M*=.07). However, in the incongruent condition participants were worse on 0x trials (0x *M*=-.04, 3x *M*=.04, and 5x *M*=.05), as indicated by a significant trial type x presentation count interaction, $F(2,68)=7.24$, $MSE=.05$, $p<.001$, $\eta_p^2=.18$. This led to very poor performance for older adults on incongruent 0x trials, (*M*=-.08), where older adults significantly *decreased* their proportion correct from forced to free report, as a 1-sample t-test showed changes in proportion correct from forced to free report in that condition to be significantly less than 0, $t(17) = 2.64$, $p<.05$. Such a decrease, where participants actually decrease their proportion correct when given the opportunity to volunteer or withhold responses has never before been shown in the metacognition literature, though a trend was detected by

Rogers et al. (in prep, Experiment 2). Koriat and Goldsmith (1996, Experiment 2) used deceptive general-knowledge questions (e.g., What is the capital of Australia?) that had devastating effects on monitoring and free report relative to control items, but the change from forced to free report accuracy was still positive. Kelley and Sahakyan (2003) found that older adults showed poorer gains from forced to free report than did young adults on deceptive items in a memory task, but, again, the gains were positive nonetheless.

Table 6. *Changes in proportion correct from forced to free report for congruent and incongruent trials in Experiment 1.*

	Presentation Count							
	0x		3x		5x		Overall	
p(Hit|Vol)-p(Hit)	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent								
Young	.08	.02	.07	.02	.07	.02	.07	.02
Older	.05	.02	.02	.02	.07	.02	.05	.02
Incongruent								
Young	.01	.03	.08	.02	.10	.02	.06	.02
Older	-.08*	.03	.00	.02	.00	.02	-.03	.02

Note: * significantly below zero, $p<.05$.

Summary

The results from Experiment 1 indicate that repeated presentations did nothing to increase the effects of context observed in Rogers et al. (in prep). This result suggests that the semantic association between the cue and the target is responsible for the effects of context observed. Context effects leading to false hearing were strongest when the cue-target pair was never presented during training. Strikingly, older adults decreased their proportion correct from forced to free report on incongruent 0x trials, which had not been shown before in the metacognition literature.

Repeated presentations did have an effect on the subjective experience of hearing, but not always in the hypothesized direction. In fact, the finding of a significant trial type x presentation count interaction in confidence, calibration bias, and changes in accuracy from forced to free report measures suggest that with repeated presentations participants were more likely to shift towards sensory information as a basis for confidence. Age groups did not differ in this shift towards sensory information with repeated presentations, as the interactions between age, trial type, and presentation count were never significant. This shift could have arisen from an expectation built during training toward hearing the sensory details of the target word that was heard during training. Violation of that expectation (e.g., presentation of an alternate word on incongruent trials), may have led to lower confidence when giving the response favored by context on 3x and 5x trials as compared to 0x trials. Further experimentation will be required to examine the nature of such an expectation. For example, repeated presentations during training may facilitate an episodic representation of the semantically associated target word (e.g., Jacoby, 1999). That episodic representation may contain sensory features of the specific utterance such as speaker information and prosody. If that information is changed during the perceptual test phase, participants may not show such a shift towards sensory information with repeated presentations.

Having found that repeated presentations did not bolster the context effects shown in Rogers et al. (in prep) came as a surprise. The results are consistent with the notion that context effects in this paradigm reflect semantic priming, and that repeated presentations during a training phase may actually reduce the likelihood of false hearing. The reverse pattern was expected; that repeated presentations would build stronger context effects and would lead to greater false hearing. If that pattern had been obtained, then in Experiment 2 a familiarization phase would have been included in the procedure. However, the results from Experiment 1 were conclusive in that such a familiarization phase is not only unnecessary, but should be avoided if one wants to test for false hearing in a situation where context effects are maximally strong.

Chapter 4 – Methods for Experiment 2

I turn now to Experiment 2, a first pass at training reductions in false hearing.

Materials and Participants

Participants. Seventy-two undergraduate and graduate students were recruited through the Washington University participant pool in St. Louis, MO. They received either $20 or course credit for their participation. These younger participants ranged in age from 18-25 years old (*M*=20.46, *SD*= 1.83). Seventy-two older adults were recruited through the Washington University Older Adult subject pool and Washington University's Volunteers for Health pool. These older participants ranged in age from 65-89 years old (*M*=73.31, *SD*= 7.21). These participants were volunteers from the St. Louis community and also received $20 for their participation. Vocabulary scores from the Shipley Institute of Living Scale (Shipley, 1946) were obtained from each participant at the end of the session to assess vocabulary proficiency. Older adults had better performance on the Shipley scale (Young *M*= 33.43, Older *M*= 34.63), $t(142) = 2.12$, $p<.05$.

Materials and Design. Experiment 2 utilized a pretest-posttest design. A set of training blocks were performed between the pretest and posttest. An equal number of young and older adults were randomly assigned to one of three training groups. The Signal Variability group (SV) received variable levels of noise masking during training. The Context Variability group (CV) received blocks that varied in terms of the ratio of congruent to incongruent trials. The No Variability group (NV), served as the control training group and received no variability during training. The manner in which the variability was implemented in the SV and CV training groups will be explained in the Procedure section.

There were very few changes made in materials from Experiment 1 to Experiment 2. All materials again consisted of 3-word sets generated for congruent and incongruent trials, and 2-word sets that were generated for baseline trials. There were only two differences in materials: a) Experiment 2 used a greater quantity of word sets and b) these word sets were broken into more component sublists, to accommodate the different portions of the experiment. In all, 144 3-word sets were compiled to serve as the congruent and incongruent trials in Experiment 2. Seventy-two 2-word sets were compiled for baseline trials.

The sublists were broken down in the following way for the context items: for the pretest and posttest phases of Experiment 2, 60 total context items were required (15 congruent pretest, 15 congruent posttest, 15 incongruent pretest, and 15 incongruent posttest). Sixty 3-word sets were then broken into four 15-set sublists, which were balanced for cue, target, and alternate word frequency, phonological confusability between target and alternate, as well as the forward and backward association strengths between the cue and the target. Those four lists served equally as often in each of the aforementioned pretest and posttest conditions, and were not used during the training-with-feedback blocks of the experiment.

For the training-with-feedback blocks, the sublist construction for the CV training group was different from NV and SV training groups. For the NV and SV training groups, 80 total context items were required (40 congruent and 40 incongruent, divided by four training blocks). The 80 context 3-word sets were then broken into eight lists of

10, balanced along the lexical and associative characteristics mentioned in Chapter 2, and then rotated across conditions equally. For the CV training group, the proportion of congruent trials to incongruent trials in each block (later referred to as context validity) was manipulated such that four 14-set sublists and four 7-set sublists were constructed. This was done to create the blocks where context was mostly valid (e.g., 14 congruent trials and 7 incongruent trials) and when context was mostly invalid (e.g., 7 congruent trials and 14 incongruent trials).   These lists were balanced and then rotated across the following congruent and incongruent conditions for each block. For all training groups, the remaining four 3-word sets were used for practice trials that were administered before the pretest.

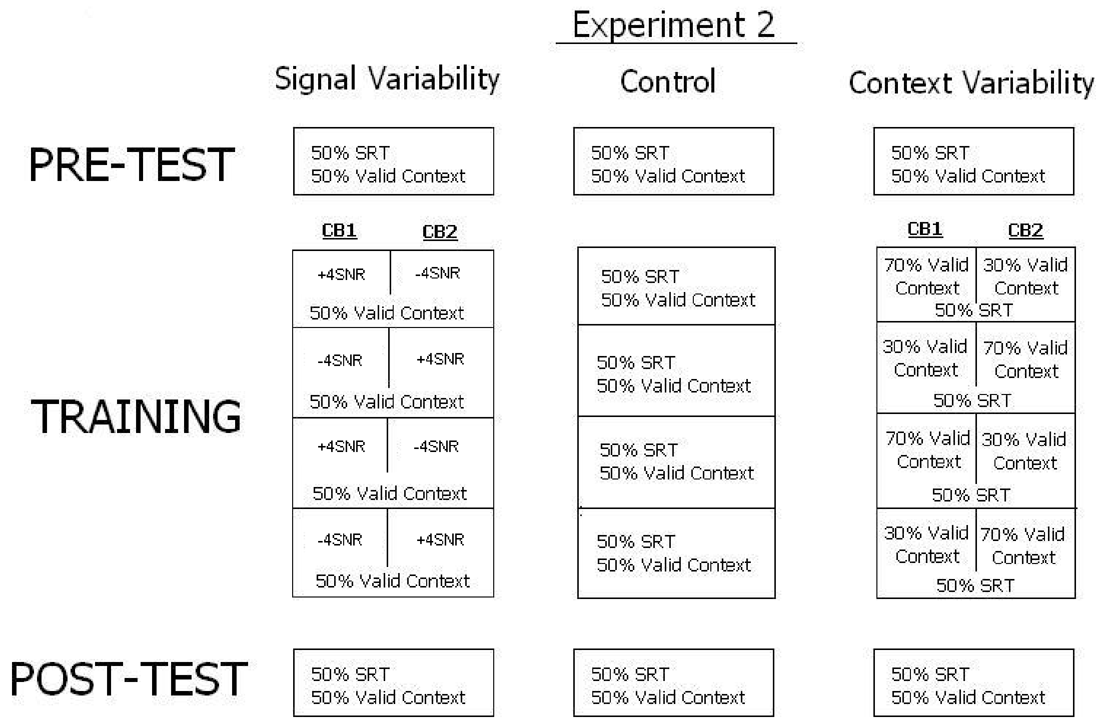
There were six sublists of baseline items. Two 15-set sublists were balanced and used for the pretest and posttest blocks—occurring equally often in each block. The baseline sets used for the training blocks were broken down into four 10-set sublists that were balanced and rotated across blocks. The remaining two 2-word sets were used for practice trials.

The sublists for the training-with-feedback blocks were rotated in the following manner. Rather than each list serving in each block equally often, training blocks one and three and training blocks two and four were yoked such that sublists served in blocks one and three equally often, or in blocks two and four equally often. Thus, items that appeared in block one appeared in block three just as often, but never appeared in blocks two or four. Yoking in this way minimized the number of rotations required for a fully balanced design. In all, eight versions of each training group were required so that each word set served equally often as a congruent and incongruent item in blocks one or three, or as a congruent and incongruent item in blocks two and four.

Procedure

The procedural differences between Experiment 1 and Experiment 2 were substantial. First, based on the results of Experiment 1, in Experiment 2 there was no familiarization phase. Second, there were six blocks of open-set cued-identification trials for each participant; a pretest block, 4 training-with-feedback blocks, and a posttest block. Third, participants were randomly assigned to one of the three training conditions: SV, CV, or NV. Each training group consisted of 24 older and 24 young adult participants. The SNR used during perceptual identification was titrated for each participant as in Experiment 1. The procedure used for titration was identical to the one used in Experiment 1.

Figure 2. *Design and procedural schematic for Experiment 2.*



Note: SNR – signal-to-noise ratio, SRT – speech reception threshold, CB – counterbalanced version of experiment.

As the procedure for Experiment 2 was more complicated than in Experiment 1, a design schematic is presented in Figure 2. After receiving instructions, participants completed six practice trials, and then began the pretest block. The pretest and posttest phases were both 45-trial blocks comprised of 15 congruent, baseline, and incongruent trials each. The noise level was set to the obtained 50% speech reception threshold (SRT). In Figure 2, the term “50% valid context” is used to refer the fact that there were an equal number of congruent and incongruent trials in the pretest and posttest, which becomes important for the CV training group, where that proportion is manipulated during training.

Following the pretest, all participants completed four training blocks that included feedback. Participants received accuracy feedback after every trial, regardless of whether the item was volunteered. After the participant chose to volunteer or withhold a trial, one of three messages appeared:

1) “Correct! +1” in green ink,

2) “Incorrect! -1” in red ink,

3) “You chose to pass on this trial,” in white ink.

Beneath one of the above messages, the correct word appeared on the screen, followed by the participant’s current score. If the participant’s score was below zero, it was still displayed. Participants’ score reset to zero after each block. At the end of each block, participants saw their scores from prior blocks, and were encouraged to beat their previous scores.

Participants in the variability training groups (e.g., SV and CV) had additional features included as part of their training. Participants in the SV training group were given variability in terms of signal quality as part of their training. There were two levels of signal quality: low, which was 4 SNR below each participant's 50% SRT, and high, which was 4 SNR above the 50% SRT. Participants completed alternating blocks during training at both low and high signal quality conditions. Whether participants completed the high signal quality trials during blocks 1 and 3 or during blocks 2 and 4 was counterbalanced across participants. Note that this counterbalance combined with the yoking of sublists mentioned in the materials section allows for each sublist to be tested at both levels of signal quality (the same was true for the variability manipulation used for CV participants as well). Participants in the CV training group were given variability in terms of context validity. In the CV group, there were two different relative proportions of congruent and incongruent trials—30% congruent/70% incongruent (low context validity) and 70% congruent/30% incongruent (high context validity). Whether participants completed the high context validity blocks first or second was counterbalanced across participants. Participants assigned to the NV group had no changes from block to block aside from the actual word sets themselves.

After completing the fourth training-with-feedback block, all participants completed the posttest block. Like the pretest block, during the posttest the noise level was set to the 50% SRT and there were 15 congruent, incongruent, and baseline trials each.

Chapter 5: Experiment 2 Results and Discussion

In Experiment 2, there were two major goals: the first was to examine whether a variability-with-feedback training procedure (CV, SV) reduced false hearing more than a practice-with-feedback control condition (NV). The second goal was to test for age differences in sensitivity to the variability manipulations present during training. In both the CV and SV training conditions, one of the two qualitatively different bases for responding (sensory and context information) was set to vary with the other remaining constant. In the SV training condition, participants were exposed to two different noise levels: 4 signal-to-noise ratio (SNR) below the titrated 50% speech reception threshold (SRT) and 4 SNR above that threshold. In the CV training condition, participants completed training blocks where the proportion of congruent trials relative to incongruent trials was varied. On half of the CV training blocks, context was valid (i.e., congruent) on 70% of the non-baseline trials. On the remaining training blocks, context was valid on 30% of non-baseline trials. I proposed in the introduction that providing this variability at training should help participants better constrain their responding to sensory information or be more flexible in their bases for responding.

This results chapter will be broken down into three major sections. The first section is concerned with pretest results and training effects. The second and third sections are concerned with age differences in sensitivity to variability. An analysis of pretest results was needed to show that random assignment of training groups was largely successful in ensuring equal pretest scores in each training group (though one unanticipated significant effect of training group was found). The analyses were also necessary to confirm that age differences in false hearing were still present when using the 0x (e.g., no familiarization phase) procedure advocated by the results of Experiment 1. Except when inconsistent with the results of Experiment 1, these results are mentioned only briefly. Lastly, training effects are discussed measure by measure, with analyses limited to hits and false alarms, confidence, calibration, monitoring, and two composite measures. These two composite variables were created to assess the extent to which participants used sensory or contextual bases for responding in the incongruent condition and will be later described.

Following the pretest and training analyses, the next two sections of this chapter are concerned with age differences in sensitivity to the variability manipulations present during training. For participants assigned to the CV training condition, age differences in hit rates, incongruent false alarms, confidence, calibration, monitoring, and composites were examined at different levels of context validity. In the SV condition, age differences were examined at different levels of sensory signal quality. These age differences were compared to those found in the NV condition, where context validity and sensory signal quality were consistent throughout training. Both the CV and SV groups will be compared to the NV group separately, resulting in two sections.

Pretest and Training Analyses

Pretest only analyses. As mentioned earlier, analyses were conducted across measures on pretest only to serve two goals. The first goal was to ensure that random assignment of young and older adults to the three training groups was successful in equating performance at pretest. Second, it was necessary to confirm that the prior findings in Experiment 1 were replicated when testing for false hearing without a familiarization phase. First, the issue of random assignment of training group is examined.

Training group differences at pretest. Random assignment of training group was largely successful in ensuring equal pretest performance across groups. Only one measure produced a significant effect of training group. For baseline monitoring, as measured by confidence discriminability, there was a small but unexpected significant main effect of training group, $F(2, 138) = 3.29$, $MSE = 574.51$, $p<.05$, $\eta_p^2 =.05$. Post-hoc tests using the Bonferroni error rate adjustment revealed that monitoring was significantly better in the CV group (*M*=14.63) than the NV (*M*=7.88) group. No other main effect or interactions involving the training group variable were significant across all pretest measures.

Age group differences at pretest. The issue of replicating the results of Experiment 1 was considered. As before in Experiment 1, baseline trials were examined separately from congruent and incongruent trials. There were only two situations in which the patterns from Experiment 1 were not replicated. First, across training groups older adults had slightly lower pretest baseline hit rates than young adults (Young *M* = .45, Older *M* = .39), suggesting the titration procedure was not as successful as in Experiment 1 in controlling for age-related differences in hearing ability. The age group difference in baseline hit rates was detected by a 2 (Age: Young, Older) X 3 (Training Group: SV, CV, NV) univariate ANOVA on baseline hit rates with a significant main effect of age, $F(1, 138) = 8.40$, $MSE = .11$, $p<.01$, $\eta_p^2 =.06$.

Despite their lower baseline hit rate, older adults were more confident in their baseline hits than were young adults (Young *M* = 50, Older *M* = 58), as revealed by a univariate ANOVA with a significant main effect of age, $F(1, 138) = 4.45$, $MSE =$ 1800.45, $p<.05$, $\eta_p^2 =.03$. This finding was different from what was previously reported in Experiment 1 and other experiments reported by Rogers et al. (in prep). Those experiments found age-group equivalence in baseline confidence. Experiment 2 differed procedurally from Experiment 1 and the experiments in Rogers et al. (in prep) in that there was no familiarization phase. The lack of familiarization could have somehow changed the nature of baseline trials for older adults. For example, in prior studies the familiarization phase could have reduced older adults' confidence in the baseline trials because they were the only trials in which the cue word had not been previously presented. Further experimentation to examine this discrepancy is warranted.

This unexpected age group difference in baseline confidence led to differences in calibration bias and monitoring. Older adults were more overconfident than young adults on pretest baseline trials, (Young $M = 0.33$, Older $M = 9.60$), $F(1, 138) = 9.77$, $MSE = 3094.88$, $p<.01$, $\eta_p^2 =.07$. This was most likely due to older adults' greater confidence in their baseline hits, as older and young adults did not significantly differ in their confidence ratings for incorrect baseline responses (Young M=39.21, Older M=41.20), $t(142)<1$, *ns*. In turn, this pattern of older adults being more confident in their hits than young adults but showing no difference for age groups in confidence for incorrect responses led to older adults having better monitoring than young adults, as measured by confidence discriminability, (Young $M = 9.36$, Older $M = 14.00$), $F(1, 138) = 4.45$, $MSE = 776.51$, $p<.05$, $\eta_p^2 =.03$.

With the exception for the findings related to the baseline condition above, the established pattern of greater or equal performance for older than young adults on congruent trials, with poorer performance for older than young adults on incongruent trials was obtained. As these findings are merely replications of patterns shown in Experiment 1 and elsewhere (Rogers et al., in prep), they are mentioned only briefly. For hit rates, older adults were equivalent to young adults on congruent trials (Young *M* = .82, Older *M* = .81), but had fewer hits than young adults on incongruent trials (Young *M* = .46, Older *M* = .35), which led to a significant interaction of trial type and age, $F(1, 138) = 10.55$, $MSE = .205$, $p<.001$, $\eta_p^2 =.07$. For false alarm rates, older adults produced more incongruent false alarms than did young adults (Young *M* = .31, Older *M* = .45), as an ANOVA on incongruent false alarm rates revealed a significant main effect of age, $F(1, 138) = 27.72$, $MSE = .774$, $p<.001$, $\eta_p^2 =.17$. Older adults were also more likely than young adults to volunteer those false alarms (Young *M* = .23, Older *M* = .42), $F(1, 138) = 51.22$, $MSE = 1.33$, $p<.001$, $\eta_p^2 =.27$.

For confidence, older adults were more confident than young adults in their congruent hits (Young *M* = 81.79, Older *M* = 92.96), as well as their incongruent false alarms (Young *M* = 69.32, Older *M* = 86.14). Young adults' confidence differed between the two responses to a greater extent than older adults', which led to a significant interaction of age and trial type, $F(1, 136) = 9.19$, $MSE = 567.84$, $p<.01$, $\eta_p^2 =.06$.

A similar pattern was observed for calibration bias, where older adults were more overconfident than young adults on congruent trials (Young *M* = -5.59, Older *M* = 3.64), but the age group difference was much greater on incongruent trials (Young *M* = 15.72, Older *M* = 37.82) confirmed by a significant interaction of age and trial type, $F(1, 138) = 17.91$, $MSE = 2982.43$, $p<.001$, $\eta_p^2 =.12$.

Older adults had better monitoring than young adults on congruent trials (Young *M* = 28.77, Older *M* = 43.40), but worse monitoring than young adults on incongruent trials (Young *M* = -4.17, Older *M* = -16.74), which led to a significant age x trial type interaction, $F(1, 128) = 19.25$, $MSE = 12491.67$, $p<.001$, $\eta_p^2 =.13$. On incongruent trials, older adults' confidence discriminability was significantly below zero, 1-sample $t(66)=7.44$, $p<.01$. Young adults' confidence discrimability on incongruent trials only trended toward differing from zero, 1-sample $t(67)=1.59$, $p<.11$. Though covered in brief, these age group differences indicate that even without a familiarization phase, older adults were more likely than young adults to hear on the basis of contextual information, leading to false hearing.

Training effects. Training effects were assessed using repeated-measures ANOVA with pretest and posttest scores added as different levels of the within-subjects variable called Test. Repeated-measures ANOVA has been criticized for use in pretest-posttest designs (Dimitrov & Rumrill, 2004; Huck & McLean, 1975; Jennings, 1988). Briefly, the two main criticisms for repeated measures ANOVA are that it is very conservative (because the between-subjects *F* term is too small, Huck & McLean, 1975), and has been frequently associated with confusion on the part of researchers in terms of interpretation of the main effect of Test. For example, the main effect of Test does not indicate a significant benefit of variability training over of the no-variability control group. Such evidence must take the form of a significant Test x Training Group interaction, revealing that performance gain at posttest was greater for one of the training groups over another. Rather, in this experiment the main effect of Test merely indicates an increase in performance shared by all training groups, and likely indicates a performance benefit resulting from practice with the task given feedback. This main effect of Test is informative only in terms of practice effects, and is reported here when significant in all training analyses.

Other statistical methods have been advocated for assessing training effects in pretest-posttest designs (Dimitrov & Rumrill, 2004; Van Breukelen, 2006). Among these are ANOVA of gain scores (subtracting the pretest score from the posttest score), ANCOVA of gain scores with pretest score added as a covariate, and multiple regression of gain scores with pretest as a predictor variable (which yields the same significance tests as ANCOVA with pretest as covariate, Cohen & Cohen, 1983, pp.417-420). Analyses were conducted using all three of these alternative methods in addition to the repeated-measures method. The results of these analyses are not reported because the conclusions reached regarding Training Group differences from those analyses did not differ from those using the repeated-measures method.

With regard to age-related differences, the ANCOVA and multiple regression procedures require an assumption of random assignment to groups (Miller & Chapman, 2001), which is not met by my manipulation of age. This assumption is most egregiously violated when strong age group differences were present at pretest (e.g., incongruent false alarms). In this case, ANOVA on gain scores could be conducted to identify if age groups differed in terms of overall training benefit from pretest to posttest. However, these differences in training benefit are more often attributable to differences before training rather than after training. As such, when reporting age-related differences in training revealed by repeated-measures ANOVA, which come in the form of Age x Test interactions, I follow Baron & Surdy (1990) in describing the pattern as age-related differences being diminished at posttest.

Baseline hit rates. The baseline hit rates during the pretest and posttest of Experiment 2 are presented in Table 7. Examination of Table 7 shows that older adults had lower baseline hit rates than young adults at pretest, but that age differences were attenuated at the posttest. Also, baseline hit rates at posttest were generally greater than those at pretest. A 2 (Age group: Young, Older) X 3 (Training Group: NV, SV, CV) X 2 (Test: Pretest, Posttest) mixed-model repeated measures ANOVA on baseline hit rates revealed a significant main effect of test, $F(1, 138) = 121.56$, $MSE = 1.31$, $p<.001$, $\eta_p^2 =.47$, as well as a significantly main effect of age, $F(1, 138) = 4.48$, $MSE = .07$, $p<.05$, $\eta_p^2 =.03$. The age x test interaction, which explains the pattern of reduced age-related differences at posttest, only approached significance, $F(1, 138) = 3.45$, $MSE = .04$, $p<.07$, $\eta_p^2 =.02$. The test x training group interaction was not significant, $F<1$, *ns*. This result suggests that neither variability training procedure (e.g., CV, SV) showed an advantage over the NV group in increasing participants' sensory-based responding.

Table 7. *Baseline hit rates at pretest and posttest in Experiment 2.*

	CV		NV		SV		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Young-Pre	0.45	0.1	0.46	0.1	0.44	0.1	0.45	0.1
Older-Pre	0.38	0.1	0.37	0.1	0.43	0.1	0.39	0.1
Young-Post	0.55	0.1	0.55	0.1	0.58	0.1	0.56	0.1
Older-Post	0.53	0.1	0.55	0.1	0.58	0.1	0.55	0.1
Young- Change	0.10	0.03	0.09	0.03	0.14	0.03	0.11	0.03
Older- Change	0.15	0.03	0.18	0.03	0.15	0.03	0.16	0.03

Table 8. *Incongruent false alarm and volunteered incongruent false alarm rates during pretest and posttest in Experiment 2.*

	CV		NV		SV		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Incongruent FAs								
Young-Pre	0.29	0.03	0.32	0.03	0.32	0.04	0.31	0.03
Older-Pre	0.49	0.04	0.43	0.04	0.43	0.04	0.45	0.04
Young-Post	0.23	0.02	0.26	0.03	0.21	0.03	0.23	0.03
Older-Post	0.37	0.04	0.29	0.02	0.29	0.03	0.32	0.03
Young- Change	-0.06	0.04	-0.06	0.04	-0.11	0.04	-0.08	0.04
Older- Change	0.14	0.04	0.03	0.04	0.08	0.04	0.09	0.04
Volunteered Incongruent FAs								
Young-Pre	0.22	0.03	0.25	0.03	0.21	0.03	0.23	0.02
Older-Pre	0.44	0.03	0.4	0.03	0.41	0.03	0.42	0.02
Young-Post	0.17	0.03	0.2	0.03	0.15	0.03	0.17	0.02
Older-Post	0.33	0.03	0.27	0.03	0.25	0.03	0.28	0.02
Young- Change	-0.05	0.04	-0.05	0.04	-0.06	0.04	-0.06	0.04
Older- Change	0.16	0.04	0.07	0.04	0.1	0.04	0.11	0.04

Incongruent false alarms and volunteered incongruent false alarms. Recall that the incongruent false alarm measure assesses the extent to which participants incorrectly report the item favored by context in the incongruent condition. The top section of Table 8 displays the rates of incongruent false alarms committed at pretest and posttest for each of the six between-subjects conditions, as well as collapsed across each of the training conditions (overall). The results show that all groups reduced their rates of incongruent false alarms at posttest relative to scores on the pretest, as revealed by a 2 (Age group: Young, Older) X 3 (Training Group: NV, SV, CV) X 2 (Test: Pretest, Posttest) mixed-model repeated measures ANOVA with a significant main effect of test, $F(1, 138) = 38.11$, $MSE = .78$, $p<.001$, $\eta_p^2 =.22$. Overall, older adults were more likely to commit incongruent false alarms than young adults, $F(1, 138) = 37.17$, $MSE = .98$, $p<.001$, $\eta_p^2 =.21$. Age-related differences trended towards being smaller at posttest than at prettest, as the age x test interaction approached significance, $F(1, 138) = 3.21$, $MSE = .07$, $p<.08$, $\eta_p^2 =.02$. No main effect or interaction involving training group was significant.

In terms of face validity, volunteered incongruent false alarms may be the best measure of false hearing because they take into account when someone responds incorrectly on the basis of context with a high degree of confidence. These data are presented in the bottom section of Table 8. Note that these rates were always smaller than the incongruent false alarm rates presented in the top section of Table 8, because they are a subset of those items. The results show that all groups reduced their rate of volunteered incongruent false alarms from pretest to posttest, as revealed by a significant effect of test, $F(1, 138) = 37.77$, $MSE = .64$, $p<.001$, $\eta_p^2 =.20$. Older adults also had higher rates of incongruent false alarms than young adults, $F(1, 138) = 66.60$, MSE = 1.63, $p<.001$, $\eta_p^2 =.33$. Age-related differences were smaller at posttest than at pretest, as indicated by a significant interaction of age and test, $F(1, 138) = 6.54$, $MSE = .12$, $p<.05$, $\eta_p^2 =.05$. This significant interaction supports the trend reported in the prior paragraph for incongruent false alarms. Both analyses of volunteered incongruent false alarm and incongruent false alarms overall revealed that all participants improved as a function of training, and that age-related differences were smaller at posttest than at pretest. None of the training groups were significantly different from one another, indicating that variability training was no better than the no-variability control group at reducing the rates of incongruent false alarms.

Composite measures. Rather than conducting training analyses for each confidence, calibration, and monitoring measure for each trial type, two composite variables were calculated to capture the extent of sensory-based and context-based responding in the incongruent condition. The data from incongruent trials were the most appropriate for doing this, as it was the only condition where context and sensory information favored different responses. To calculate the sensory-based responding composite, each participant's individual incongruent hit rate was averaged with mean confidence in incongruent hits. Confidence ratings were divided by 100 to equate scales with hit rate. The context-based responding composite was calculated the same way, except that the rate of incongruent false alarms was averaged with mean confidence in those responses. Confidence ratings were included in these composites to factor in how much participants thought they "heard" on the basis of sensory or contextual information. The results at pretest and posttest for both measures are presented in Table 9.

Table 9. *Composite measures for sensory-based and context-based responding during pretest and posttest in Experiment 2.*

	CV		NV		SV		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Sensory-Based								
Young-Pre	1.18	0.04	1.11	0.03	1.02	0.04	1.1	0.04
Older-Pre	0.97	0.06	1.04	0.08	1.12	0.05	1.04	0.06
Young-Post	1.13	0.05	1.11	0.05	1.16	0.04	1.13	0.05
Older-Post	0.99	0.06	1.14	0.05	1.18	0.04	1.1	0.05
Young- Change	-0.05	0.05	0.00	0.05	0.14	0.05	0.03	0.05
Older- Change	-0.14	0.05	0.03	0.05	0.02	0.05	-0.03	0.05
Context-Based								
Young-Pre	1.02	0.04	1.01	0.05	0.98	0.06	1	0.05
Older-Pre	1.32	0.05	1.29	0.04	1.35	0.04	1.32	0.04
Young-Post	0.8	0.04	0.9	0.04	0.86	0.05	0.85	0.04
Older-Post	1.08	0.06	1.11	0.04	1.03	0.07	1.07	0.06
Young- Change	-0.22	0.05	-0.11	0.05	-0.12	0.05	-0.15	0.05
Older- Change	0.28	0.05	0.21	0.05	0.17	0.05	0.22	0.05

Examination of the top panel of Table 9 shows that with regard to sensory-based responding, there was little in the way of differences between pretest and posttest. In fact, the 2 (Age group: Young, Older) X 3 (Training Group: NV, SV, CV) X 2 (Test: Pretest, Posttest) mixed-model repeated measures ANOVA on sensory composite scores revealed only a marginally significant main effect of test, F(1, 138) = 2.99, MSE = .03, p<.09, η_p^2 =.02. The only significant effect revealed by the ANOVA was a significant interaction of training group and age, F(1, 138) = 4.27, MSE = .09, p<.05, η_p^2 =.06. Post-hoc analyses collapsing across pretest and posttest revealed that older adults in the CV condition had significantly lower sensory composite scores than older adults in the SV condition, F(2, 138) = 4.49, MSE = .05, p<.05, η_p^2 =.06, and their young CV counterparts, F(1, 138) = 8.98, MSE = .09, p<.01, η_p^2 =.06.

Unlike the results for the sensory composite, the bottom panel of Table 9 shows that there was considerable change from pretest to posttest on the context composite measure. Participants had lower scores on the context composite on the posttest than on the pretest, as indicated by a significant main effect of test, F(1, 138) = 57.28, MSE = .82, p<.001, η_p^2 =.29. Across training groups, older adults were more likely to respond on the basis of contextual information than young adults, as indicated by a significant main effect of age, F(1, 138) = 68.16, MSE = 1.58, p<.001, η_p^2 =.33. The interaction between age and test was not significant, F<1, *ns*, indicating that age-related differences at posttest were similar to those at pretest.

The results of the sensory and context composite analyses are in line with pattern of results given by the training analyses of baseline hits and incongruent false alarms. It seems that the most substantial effects of training were on reducing the amount of context-based responding. Training did not show substantial increases in sensory-based responding, except in the case of older adults' baseline hits. While age-related differences in baseline hits were smaller at posttest than at pretest, those age-related differences at pretest more likely reflected a flawed titration procedure than true age-related differences. Prior studies using this procedure found age equivalence in baseline hits rates (Rogers, et al., in prep).

The results allow the conclusion that the variability-based training procedures did not differ from the NV control group in reducing false hearing. Thus, the efficacy of such a small amount of variability training as a method for reducing false hearing is cast substantially in doubt. While there were reductions in context-based responding that occurred in all training groups from pretest to posttest, these effects most likely reflect the benefits of extended practice with this procedure. This kind of practice did not completely eradicate false hearing, and while disappointing, it does show that false hearing is a robust phenomenon, particularly in older adults. We turn now to the results regarding age-related differences in sensitivity to the variability present during training.

Sensitivity to Variability in Context Validity: CV vs. NV

While the results of the regression analyses were conclusive in showing that no training group demonstrated large advantages at posttest, the question still remained as to whether older and young adults differed in their sensitivity to the variability manipulations. Rather than comparing all three groups at the same time, the analyses were separated into two sections: comparisons between the CV group and the NV group, and comparisons between the SV group and the NV group. Handling the results in this way allowed for a simpler separation of the effects of context validity variability and signal quality variability. Unless otherwise specified, only effects found to be significant at the α=.05 significance level that were not involved in a higher-order interaction are reported. Also, to account for the age-related difference in baseline hit rate found during the pre-test, I report only effects that remained significant in a subsequent ANCOVA with baseline hit rate during the pretest added as a covariate.

Recall that the difference between the NV and CV training groups was that during training, participants in the CV condition completed alternating blocks that varied in the proportion of congruent and incongruent trials. In both the NV and CV groups, one third of all trials were baseline trials. In the NV group, 50% of the non-baseline trials were congruent and 50% were incongruent. In the CV group, there were two different relative proportions of congruent and incongruent trials—30% congruent/70% incongruent (low context validity) and 70% congruent/30% incongruent (high context validity). Comparing the CV and NV groups amounts to three different validity levels, low CV (30% valid), NV (50% valid), and high CV (70% valid). For the purposes of ANOVA, trials in the NV group were randomly coded as low context and high context validity trials to equate number of cells with the CV condition. Because high context and low context trials in the NV condition did not differ with respect to any actual manipulation, any differences obtained must be ascribed to sampling error. When reporting the NV data, the means are collapsed across these random codes. For example, in Table 10 the hit and false alarm rates for NV and CV participants during the training phases of Experiment 2 are presented. The low validity CV results are presented in the leftmost column, followed by the NV results, the high validity CV results, and the overall results collapsed across training group and validity. It is important to note that the low CV, NV, and high CV columns do not correspond to three different between-subjects conditions, but instead to two groups of participants (NV and CV), with the CV groups broken into the two context validity levels, which were manipulated within-subjects.

Table 10. *Hit and false alarm rates during the training phase for CV and NV participants in Experiment 2.*

Context Validity	CV - Low		NV		CV - High		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent Hits								
Young	0.80	0.02	0.83	0.02	0.81	0.02	0.82	0.02
Older	0.88	0.02	0.83	0.02	0.87	0.01	0.85	0.02
Baseline Hits								
Young	0.46	0.02	0.49	0.03	0.46	0.02	0.48	0.02
Older	0.43	0.03	0.45	0.03	0.41	0.02	0.43	0.03
Incongruent Hits								
Young	0.38	0.03	0.41	0.03	0.38	0.03	0.40	0.03
Older	0.29	0.02	0.28	0.03	0.30	0.03	0.29	0.03
Incongruent FAs								
Young	0.29	0.02	0.27	0.02	0.37	0.03	0.30	0.02
Older	0.42	0.02	0.42	0.03	0.41	0.03	0.42	0.03

Hits and false alarms. For hit rates during the training phase, the pattern was much the same as shown during the prettest, though here baseline trials were included in the analyses. Table 10 shows that older adults had greater hit rates than young adults on congruent trials (Young *M* = .82, Older *M* = .85), lower hit rates on baseline trials (Young *M* = .48, Older *M* = .43), and even lower hit rates on incongruent trials (Young *M* = .40, Older *M* = .29), as a 3 (Trial type: Congruent, Baseline, Incongruent) X 2 (Context Validity: Low, High) X 2 (Training Group: NV, CV) X 2 (Age: Young, Older) repeated-measures mixed-model ANOVA on hit rates revealed a significant age x trial type interaction, $F(2, 184) = 19.55$, $MSE = .25$, $p<.001$, $\eta_p^2 =.18$. There was no significant effect of context validity or training group on hit rates, *F's*<1, *ns*.

As expected, older adults were more likely to commit incongruent false alarms than young adults (Young *M* = .30, Older *M* = .42), as a 2 (Context Validity: Low, High) X 2 (Training Group: NV, CV) X 2 (Age: Young, Older) repeated-measures mixed-model ANOVA on incongruent false alarms revealed a significant main effect of age, $F(1, 92) = 25.84$, $MSE = .67$, $p<.001$, $\eta_p^2 =.22$. Examining the incongruent false alarms rates in Table 10, it seems that while context validity did not have an effect on older adult CV participants (Older low *M* =.42, Older high *M* =.41), young adults in the CV training group were more likely to commit incongruent false alarms when context validity was high (Young low *M* =.29, Young high *M* =.37). Similar to the age differences in sensitivity to shifting payoff matrices observed by Baron and Surdy (1990), young adults may have been sensitive to the higher level of context validity, and were more likely to produce the response favored by context when context was higher in validity. Older adults may not have been sensitive to this change in context validity, and, so, responding did not change. While this 3-way interaction of age, training group, and context validity qualifying this finding was only marginally significant when NV participants were included in this analysis, $F(1,91) = 3.80$, $MSE = .04$, $p<.06$, $\eta_p^2 = .04$. A separate ANOVA on CV participants revealed the age x context validity interaction to be significant, $F(1,45) = 6.23$, $MSE = .06$, $p<.02$, $\eta_p^2 = .12$.

Confidence in hits and false alarms. After showing some evidence for age differences in sensitivity to the context validity manipulation in incongruent false alarm rates, I turn now to the results of the subjective experience measures. The mean confidence ascribed to congruent hits, baseline hits, and incongruent false alarms are presented in Table 11. Note that confidence in congruent hits and incongruent false alarms should reflect the level of confidence when providing the response favored by context, whereas confidence in baseline hits reflects responses made on the basis of sensory information. Looking at Table 11, one finds very little if any effect of context validity on the pattern of responding for young and older adults. A 3 (Trial type: Congruent, Baseline, Incongruent) X 2 (Context Validity: Low, High) X 2 (Training Group: NV, CV) X 2 (Age: Young, Older) repeated-measures mixed-model ANOVA on mean confidence in these responses found no significant effect of training group or context validity.

Comparing the top and bottom half of Table 11 does reveal the expected pattern that older adults were more likely to be confident when providing a response favored by context. Overall, older adults were more likely to be confident than young adults in their congruent hits (Young M =70.84, Older M = 82.68), and incongruent false alarms (Young M = 60.89, Older M = 76.72), but not baseline hits (Young M = 53.22, Older M = 55.76), as revealed by a significant age x trial type interaction, $F(2, 182) = 171.63$, MSE = 1866.60, $p<.001$, $\eta_p^2 =.15$. One older adult and one young adult were excluded from analysis for not committing any incongruent false alarms.

Table 11. *Mean confidence in congruent hits, baseline hits, and incongruent false alarms during the training phase for CV and NV participants in Experiment 2.*

		Congruent Hit		Baseline Hit		Incongruent FA	
Young		*M*	*SE*	*M*	*SE*	*M*	*SE*
	CV - Low	70	2	52	3	61	2
	NV	72	2	54	2	62	3
	CV - High	70	2	53	3	59	3
	Overall	71	2	53	3	61	3
Older							
	CV - Low	82	3	54	3	77	3
	NV	86	3	61	4	79	4
	CV - High	81	3	54	3	75	4
	Overall	83	3	56	3	77	4

Calibration and monitoring. I now turn to the data regarding calibration and monitoring, which are presented in the top and bottom panels of Table 12, respectively. The calibration results were consistent with the results obtained for confidence and hit rates in that there were no significant effects of training group or context validity. Looking at the top panel of Table 12, there was a tendency for older adults in the NV condition to be more overconfident across trial types, but the age x training group interaction was not significant, $F(1, 91) = 1.71$, $MSE = 1661.21$, $p<.20$, *ns*. As in prior studies, older adults were less underconfident than young adults on congruent trials (Young $M = -14.49$, Older $M = -6.60$), but more overconfident than young adults on baseline trials (Young $M = -1.53$, Older $M = 5.44$) and incongruent trials (Young $M = 17.49$, Older $M = 35.39$), as confirmed by a significant age x trial type interaction, $F(2, 184) = 8.73$, $MSE = 1424.03$, $p<.001$, $\eta_p^2 = .09$.

For monitoring, again measured by confidence discriminability, the typical pattern of older adults exhibiting better monitoring than young adults on congruent trials (Young $M = 21.81$, Older $M = 29.45$), poorer monitoring than young adults on incongruent trials (Young $M = -1.20$, Older $M = -16.96$), and equivalent monitoring for age groups on baseline trials (Young $M = 14.15$, Older $M = 15.44$) was obtained, as revealed by a significant age x trial type interaction, $F(2, 81) = 19.58$, $MSE = 5489.02$, $p<.001$, $\eta_p^2 = .20$. For these analyses, 8 older adults and 3 young adults were excluded because of either 100% performance on congruent trials or for having never providing an incongruent false alarm.

Table 12. *Calibration and monitoring data during the training phase for CV and NV participants in Experiment 2.*

CALIBRATION

Context Validity	CV - Low		NV		CV - High		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent								
Young	-14	2	-14	2	-15	2	-14	2
Older	-9	4	-2	4	-10	3	-7	4
Baseline								
Young	-3	3	-1	2	-1	3	-2	3
Older	3	3	7	4	4	3	5	3
Incongruent								
Young	17	3	15	3	18	3	17	3
Older	31	4	41	5	33	4	35	4

MONITORING

Context Validity	CV - Low		NV		CV - High		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent								
Young	23	4	21	2	22	3	22	3
Older	22	4	33	5	33	4	29	4
Baseline								
Young	15	3	13	2	15	2	14	2
Older	13	3	17	2	16	3	15	3
Incongruent								
Young	0	2	-2	2	-1	3	-1	2
Older	-24	3	-11	3	-15	4	-17	3

Despite the null effects of context validity and training group on hit rates, confidence, and calibration, effects of training group and context validity were found on monitoring for older adults. Careful examination of the bottom panel of Table 12 shows that across trial type older adults in the CV training group had poorer monitoring in the low validity condition than they did in the high validity condition. Their monitoring in the low validity condition was also poorer than that of older adults in the NV condition. This resulted in a significant age x training group x context validity interaction, $F(1, 81) = 10.84$, $MSE = 1748.74$, $p<.001$, $\eta_p^2 = .12$.

This monitoring pattern could have resulted from older adults' increased reliance on contextual information, which may have led to poorer monitoring when context was invalid on 70% of trials. When contextual information was high in validity, age groups did not differ in monitoring. Because monitoring was so poor in the low validity condition, it suggests that older adults were not as sensitive as young adults to the manipulation of context validity. If older adults were not aware of the manipulation, this could explain why older adults in the CV condition showed poorer training effects relative to the rest of the groups in Experiment 2. It is also important to note that no differences in confidence, monitoring, or calibration occurred as a function of context validity for young adults, despite results indicating that young adults were more likely to produce incongruent false alarms in the high CV condition. How older adults could have been unaware will be addressed in the General Discussion.

Composites. One final comparison involves the sensory and context composite variables introduced in the discussion of the regression analyses. The top panel of Table 13 presents the sensory-based responding composite, which showed little effect of age, training group, or context validity, except in the case of CV older adults when context validity is low. It seems that when context validity was low, older adults were less likely to respond on the basis of sensory information than in other conditions (Older low CV *M*=.84, Older NV *M*=.97, Older high CV *M*= .95). While a 2 (Age: Young, Older) x 2 (Validity: Low, High) X 2 (Training Group: NV, CV) mixed-model repeated measures ANOVA on sensory composite scores revealed no significant effects, a separate ANOVA on CV participants only did reveal a significant age x validity interaction, $F(1, 45) = 4.05$, $MSE = .09$, $p=.05$, $\eta_p^2 =.09$. This subsequent ANOVA was justified on the basis that we did not expect any effects of validity in NV participants, as the trials were randomly coded high and low validity with no actual manipulation.

While older adults were nearly as likely as young adults to respond on the basis of sensory information, older adults were more likely than young adults to respond on the basis of contextual information. The context-based responding composite results, presented in the bottom panel of Table 13, show that older adults scored higher than young adults on the context composite in all conditions (Young *M*=.92, Older *M*=1.19), as revealed by a 2 (Age: Young, Older) x 2 (Validity: Low, High) X 2 (Training Group: NV, CV) mixed-model repeated measures ANOVA with a significant effect of age, $F(1, 90) = 38.16$, $MSE = 2.77$, $p<.001$, $\eta_p^2 =.30$. One older adult and one young adult were excluded from analysis because they did not commit an incongruent false alarm. No other significant effects were found. A separate ANOVA on CV participants also revealed an absence of significant effects other than age.

Table 13. *Sensory-based and context-based responding composite for NV and CV participants during the training phase of Experiment 2.*

Context Validity	CV – Low		NV		CV - High		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Sensory-Based								
Young	0.98	0.04	1.00	0.04	0.96	0.05	0.98	0.04
Older	0.84	0.04	0.97	0.05	0.95	0.04	0.92	0.04
Context-Based								
Young	0.90	0.03	0.89	0.03	0.97	0.05	0.92	0.04
Older	1.18	0.04	1.22	0.05	1.16	0.06	1.19	0.05

While the results from incongruent false alarms, monitoring and sensory-based responding composite do seem to indicate an age-related difference in sensitivity to the context validity manipulation, there were no effects of context validity on hit rates, confidence, calibration, or on the context-based responding composite. Potentially, the validity manipulation was not strong enough. Perhaps if the relative proportion of congruent to incongruent trials was more severe in the CV condition (e.g., 90% congruent vs. 90% incongruent), the pattern of results obtained in monitoring could have also generalized to accuracy and other measures of subjective experience. Another possibility is that the presence of feedback may attenuate some effects of context validity. Given that participants were receiving trial-by-trial feedback in the CV condition, they may have been less likely to be misled by variability in context validity and show weaker context effects than they would in no-feedback conditions. However, no experiment to date has examined context variability without feedback in false hearing, so further investigation is required.

Sensitivity to Variability in Signal Quality: SV vs. NV

By comparing performance during training of SV participants to NV participants, age and training group differences in sensitivity to variation in signal quality were examined in terms of hit rates, false alarm rates, confidence, calibration, monitoring, and composite measures of sensory-based and context-based responding. Comparing the SV and NV groups amounts to three different signal levels, low SV, NV, and high SV. While signal quality was not varied in the NV condition, trials were randomly coded as low and high signal quality trials to ease comparison with the SV condition. As before in the comparison of NV participants to CV participants, any differences obtained between low and high signal quality trials for NV participants must be ascribed to sampling error, because those trials did not differ with respect to any actual manipulation. As before with the CV participants, note that SV training group columns depicted in tables and figures represent one group split on the within-subjects manipulation of signal quality, not two different SV training groups.

Hits and false alarms. The typical pattern of stronger context effects for older as compared to young adults can be observed by comparing hit rates in Table 14, where older adults show greater facilitation of context on congruent trials, but also greater interference on incongruent trials. This context effect was detected by a 3 (Trial type: Congruent, Baseline, Incongruent) X 2 (Signal Quality: Low, High) X 2 (Training Group: NV, SV) X 2 (Age: Young, Older) repeated-measures mixed-model ANOVA on hit rates which revealed a significant age x trial type interaction, $F(2, 184) = 17.65$, $MSE = .22$, $p<.001$, $\eta_p^2 =.16$. While matched in terms of baseline (Young M = .48, Older M = .48), older adults had better accuracy on congruent trials (Young M = .81, Older M = .84), and poorer accuracy on incongruent trials (Young M = .41, Older M = .32), when collapsing across training group.

Varying the levels of noise had a strong effect on hit rates for SV participants. Hit rates differed as a function of signal quality for SV participants to a greater extent in baseline (low M=.34, high M=.64) and incongruent trials (low M=.23, high M=.51), than in congruent trials (low M=.76, high M=.89), which resulted in a significant training group x signal quality x trial type interaction, $F(2, 184) = 12.14$, $MSE = .10$, $p<.001$, $\eta_p^2 =.12$. Examination of Table 14 also shows that the signal quality of NV participants yielded values intermediate to those of the high and low SV conditions. Qualified by the higher-order interaction, two separate ANOVAs were run for each training group. The ANOVA on NV participants revealed no significant interaction of signal quality and trial type, $F<1$, *ns*. The ANOVA on SV participants did reveal a significant signal quality x trial type interaction, $F(2, 90) = 5.45$, $MSE = .04$, $p<.01$, $\eta_p^2 = .11$. The diminished effect of signal quality in congruent trials could potentially have been due to a ceiling effect. However, on congruent trials sensory information was not the only valid basis for responding, and, so, a manipulation of signal quality may have affected responding less strongly than on baseline and incongruent trials. Even with poor signal quality, hit rates on congruent trials in the SV condition were still very high (.76), speaking strongly to the support provided by context.

Table 14. *Hit and false alarm during the training phase for SV and NV participants in Experiment 2.*

Signal Quality	SV - Low		NV		SV - High		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent Hits								
Young	0.73	0.02	0.83	0.02	0.86	0.02	0.81	0.02
Older	0.79	0.03	0.83	0.02	0.91	0.01	0.84	0.02
Baseline Hits								
Young	0.34	0.02	0.49	0.03	0.61	0.02	0.48	0.02
Older	0.34	0.02	0.45	0.03	0.66	0.02	0.48	0.02
Incongruent Hits								
Young	0.27	0.02	0.41	0.03	0.55	0.03	0.41	0.03
Older	0.20	0.02	0.28	0.03	0.47	0.03	0.32	0.03
Incongruent FAs								
Young	0.33	0.03	0.27	0.02	0.20	0.02	0.27	0.02
Older	0.43	0.03	0.42	0.03	0.30	0.03	0.38	0.03

The incongruent false alarm rate was also influenced by the manipulation of signal quality for SV participants, with the NV signal quality again yielding intermediate values. SV participants trended toward a higher rate of incongruent false alarms than NV participants at low signal quality conditions (NV *M*=.34, SV *M*=.38), but committed fewer incongruent false alarms than NV participants in high signal quality conditions (NV *M*=.35, SV *M*=.25), as revealed by a significant signal quality x training group interaction, $F(1, 92) = 25.88$, $MSE = .27$, $p<.001$, $\eta_p^2 = .22$. This training group x signal quality interaction did not change with age, $F<1$, *ns,* but older adults were more likely to false alarm than young adults overall (Young *M* = .27, Older *M* = .38), as indicated by a significant main effect of age, $F(1, 92) = 33.82$, $MSE = .78$, $p<.001$, $\eta_p^2 = .27$. When analyses were restricted to the rate of volunteered incongruent false alarms (not shown), the same interaction of signal quality and training group was obtained, $F(1, 92) = 15.50$, $MSE = .17$, $p<.001$, $\eta_p^2 = .14$, as well as the main effect of age, $F(1, 92) = 58.37$, $MSE = 1.33$, $p<.001$, $\eta_p^2 = .39$.

Confidence in hits and false alarms. The mean confidence ratings in congruent hits, baseline hits, and incongruent false alarms are presented separately for young and older adults in Figures 2 and 3, respectively. In both figures, SV participants' confidence at low and high signal quality levels are presented relative to that of NV participants. For ease of interpretation, NV participants' confidence is collapsed across signal quality, and listed at the 50% identification threshold. These figures show that all participants were more confident when they were providing the response favored by context (e.g., congruent hits and incongruent false alarm) than when responding on the basis of sensory information (e.g., baseline hit), reflecting a “V” pattern for confidence first described by Rogers et al. (in prep).

The figures indicate that confidence in congruent hits and incongruent false alarms change little as a function of the different noise levels, revealing a lack of sensitivity to the variability of the signal quality. That is, despite the result in Table 14 that increasing signal quality from low to high in the SV condition increased congruent hits (Young *M*=+.13, Older *M*=+.12), and decreased incongruent false alarms, (Young *M*=-.13, Older *M*=-.13), increasing signal quality had little effect in confidence in these responses. This finding is surprising given that the differences in signal quality in these conditions were substantial. Yet, both congruent hits and incongruent false alarms were the responses that were favored by context. For these responses, sensory information was less relevant. Mean confidence in baseline hits varied more substantially than did the other two responses as a function of signal quality. When sensory information was the only basis for responding, a manipulation of signal quality had greater effect. A surprising but intriguing result comes from a comparison of the changes in baseline confidence as a function of signal quality from Figure 3 to Figure 4; which compares young and older adults. That comparison shows that older adults’ confidence was more greatly influenced by changes in signal quality (low *M*=47, high *M*=70) than was young adults’ (low *M*=47, high *M*=57). When sensory information was the only available basis for responding, older adults were more sensitive than young adults to variability in signal quality.

Figure 3.

Mean confidence ratings during the training phase for young NV and SV participants in Experiment 2.

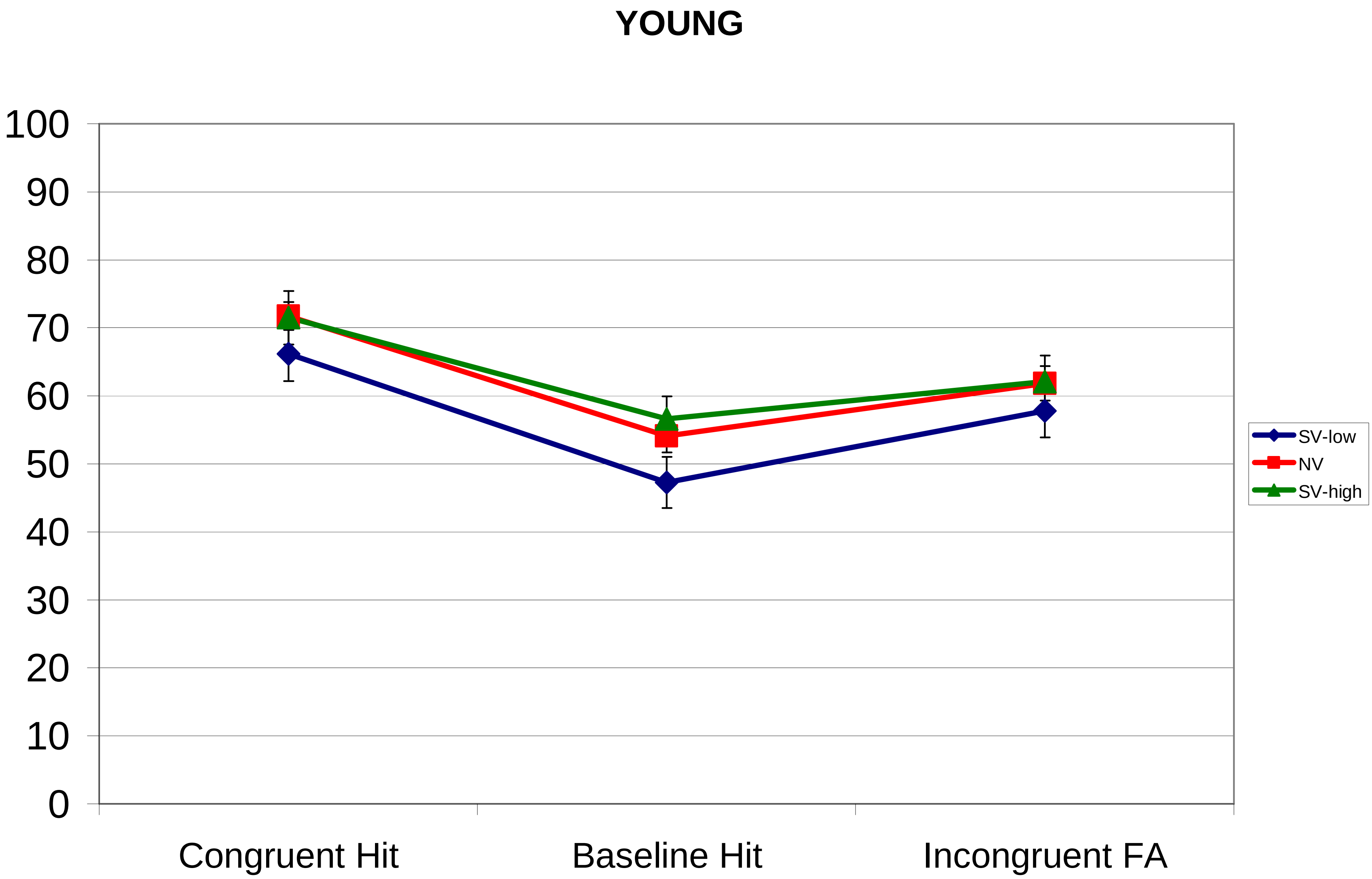
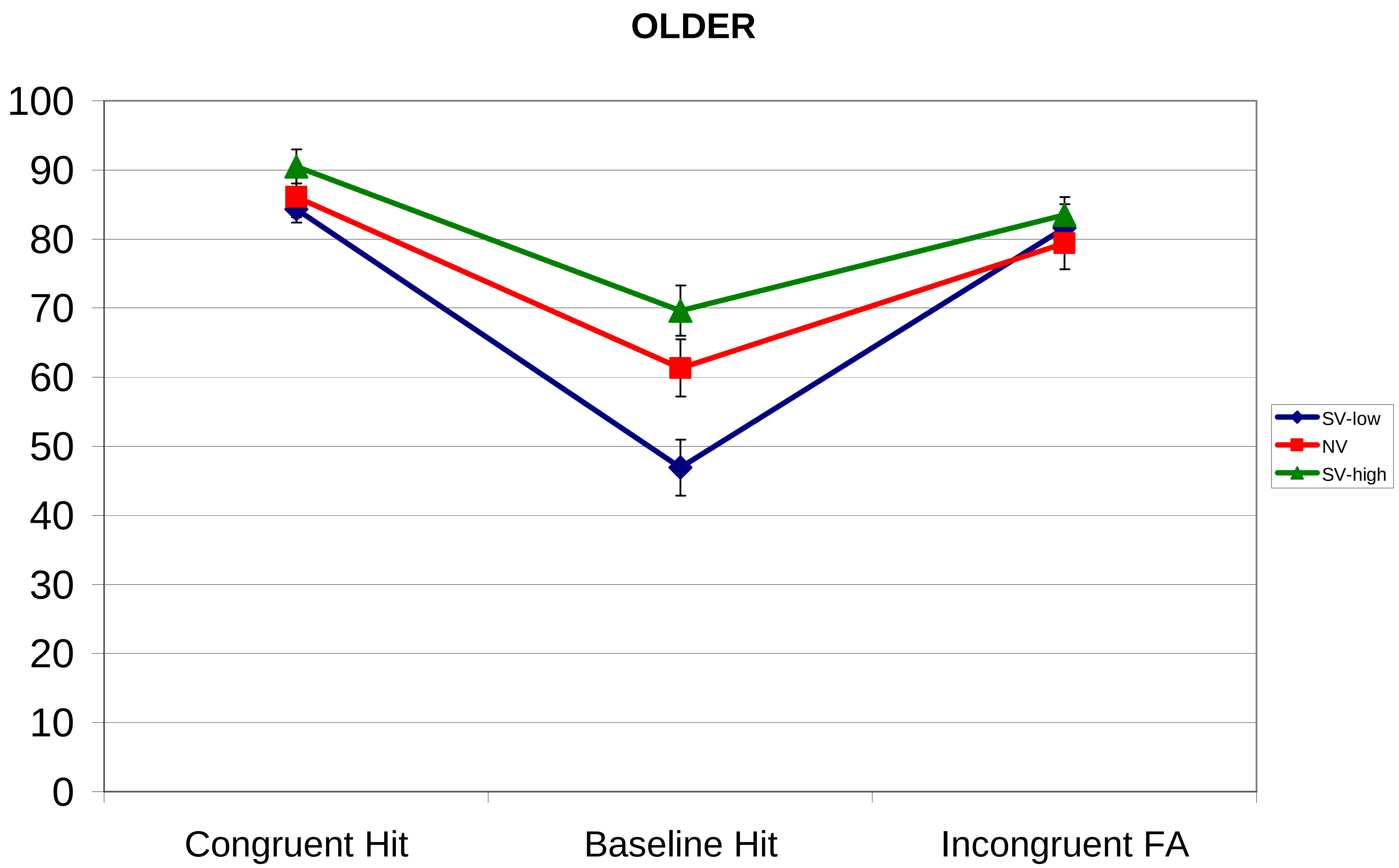


Figure 4.

Mean confidence ratings during the training phase for older NV and SV participants in Experiment 2.



Statistically, the 3 (Trial type: Congruent, Baseline, Incongruent) X 2 (Signal Quality: Low, High) X 2 (Training Group: NV, SV) X 2 (Age: Young, Older) repeated-measures mixed-model ANOVA on mean confidence in congruent hits, baseline hits, and incongruent false alarms revealed a significant 4-way interaction of training group, age, trial type, and signal quality, $F(2, 182) = 4.72$, $MSE = 206.10$, $p<.001$, $\eta_p^2 =.14$. One young adult subject in the NV condition was excluded from analysis for not having committed an incongruent false alarm. Separate ANOVAS were conducted for each training group. The 3-way interaction of age, signal quality, and trial type was significant for SV participants, $F(2, 190) = 9.19$, $MSE = 435.96$, $p<.001$, $\eta_p^2 =.17$, but not for NV participants, $F<1$, *ns.* Separate ANOVAs for each age group within the SV training group revealed a significant level x trial type interaction only for older adults, $F(2, 46) = 28.91$, $MSE = 1452.69$, $p<.001$, $\eta_p^2 =.56$. For older adults in the SV condition, simple effects tests of signal quality were significant for congruent, $F(1,23) = 19.01$, $p<.001$, $\eta_p^2 =.45$, and baseline trials, $F(1,23) = 56.69$, $p<.001$, $\eta_p^2 =.71$, but not for incongruent trials, $F<1$, *ns.*

This novel finding suggests that when context is not available as a basis for responding, older adults' confidence judgments are more sensitive than young adults to variations in signal quality. This sensitivity should be capitalized upon in future training studies. However, when providing the response favored by context, the manipulation of signal quality had little effect for either older or young adults. In those cases, participants' confidence could have been based predominantly on context information alone. Future training studies could take advantage of older adults' greater sensitivity in situations where context is not available, and then slowly add contextual information to encourage flexibility across bases for responding.

To ensure that the confidence pattern above was not simply an artifact of using conditionalized data, a 2 (Age:Young, Older) X 3 (Trial type: Congruent, Baseline, Incongruent) X 2 (Signal Quality: Low, High) mixed-model repeated-measures ANOVA was run on overall mean confidence rating for SV participants. Note that these overall mean confidence ratings are the data from which the hit rates are subtracted from to obtain calibration bias scores. As with the ANOVA run on the conditionalized confidence ratings, this ANOVA revealed a significant 3-way interaction of age, signal quality, and trial type, $F(2, 92) = 10.75$, $MSE = 4281.01$, $p<.001$, $\eta_p^2 =.19$. From low to high signal quality, age groups showed similar increments in confidence on congruent (Young *M*=+9.82, Older *M*=+12.79) and incongruent trials (Young *M*=+10.28, Older *M*=+10.34), yet older adults had larger increments in confidence on baseline trials than young adults (Young *M*=+13.23, Older *M*=+26.34). These results suggest that older and young adults' confidence ratings may have been more sensitive to the changes in signal quality than suggested by the conditionalized data on congruent and incongruent trials, but the pattern of greater sensitivity for older adults than young adults on baseline trials was again obtained.

Calibration and monitoring. The results on calibration, which are presented in the top panel of Table 15 revealed the familiar pattern that older adults were more underconfident than young adults on congruent trials (Young *M* = -4.17, Older *M* = -15.14), similarly calibrated on baseline trials (Young *M* = -1.79, Older *M* = 3.24), and much more overconfident than young adults on incongruent trials (Young *M* = 13.35, Older *M* = 35.37), which resulted in a significant interaction of trial type and age, $F(2, 184) = 20.57$, $MSE = 3394.39$, $p<.001$, $\eta_p^2 =.18$. This pattern of results was consistent with the calibration results obtained in Experiment 1, where older adults showed greater calibration bias on congruent and incongruent trials because of their greater reliance on contextual information as a basis for responding.

Participants in the SV condition were not fully sensitive to the increases in baseline and incongruent accuracy that came as a function of improving signal quality. Comparing the leftmost and rightmost columns of the top panel of Table 15 shows that as signal quality improves for SV participants, both young and older adults become more underconfident on baseline and incongruent trials. This shift improves calibration on incongruent trials. This shift toward underconfidence for SV participants on baseline (low M=4.20, high M=-5.71) and incongruent trials (low M=30.80, high M=13.20) was age invariant, and resulted in a significant signal quality x training group x trial type interaction, $F(2, 184) = 9.25$, $MSE = 801.05$, $p<.001$, $\eta_p^2 =.09$. A separate ANOVA for SV participants revealed a significant, signal quality x trial type interaction, $F(2,92) = 19.23$, $MSE = 1576.37$, $p<.001$, $\eta_p^2 =.71$. The separate ANOVA for NV participants did not reveal a significant signal quality x trial type interaction, $F<1$, *ns*.

Table 15. *Calibration and monitoring data during the training phase for CV and NV participants in Experiment 2.*

CALIBRATION

Signal Quality		SV - Low		NV		SV - High		Overall	
		M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent									
	Young	-14	3	-14	2	-18	4	-15	3
	Older	-5	2	-2	4	-4	2	-4	3
Baseline									
	Young	4	3	-1	2	-9	4	-2	3
	Older	4	3	7	4	-2	4	3	4
Incongruent									
	Young	21	4	15	3	3	4	13	4
	Older	40	4	41	5	23	5	35	5

MONITORING

Signal Quality		SV - Low		NV		SV - High		Overall	
		M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Congruent									
	Young	28	3	21	2	24	3	24	3
	Older	52	4	33	5	32	5	39	5
Baseline									
	Young	15	2	13	2	14	3	14	2
	Older	15	2	17	2	14	3	15	2
Incongruent									
	Young	-4	2	-2	2	-1	3	-2	2
	Older	-36	5	-11	3	-15	3	-21	4

The pattern of results for monitoring, shown in the bottom panel of Table 15, shows that older adults had better monitoring on congruent trials, equivalent monitoring on baseline trials, and poorer monitoring on incongruent trials than young adults as indicated by a significant age x trial type interaction, $F(2, 154) = 32.97$, $MSE = 9386.23$, $p<.001$, $\eta_p^2 = .30$. Collapsed across signal quality conditions, this trial type x age interaction was stronger for SV participants, such that older SV participants' monitoring was greater than that of older NV participants on congruent trials (Older SV $M = 42.10$, Older NV = 33.25), and poorer than older NV participants on incongruent trials (Older SV $M = -25.49$, Older NV $M = -10.64$). This training group effect was selective to older adults, resulting in a significant trial type x age x training group interaction, $F(2, 154) = 3.25$, $MSE = 922.59$, $p<.05$, $\eta_p^2 = .04$. Because monitoring was assessed using the confidence discriminability measure, 14 participants (5 young and 9 older adults) had to be excluded from analysis for either not providing any incongruent false alarms or for not having any incorrect responses on congruent trials.

When signal quality was improved, SV older adults showed diminished context effects on both congruent (Older low $M = 52.24$, Older high $M = 32.19$) and incongruent trials (Older low $M = -36.35$, Older high $M = -14.79$). Young adults did not show this effect on congruent (Young low $M = 27.51$, Young high $M = 24.63$) or incongruent trials (Young low $M = -4.44$, Young high $M = -1.34$), which resulted in a significant age x trial type x signal quality interaction, $F(2, 154) = 5.53$, $MSE = 1067.88$, $p<.01$, $\eta_p^2 = .07$. This finding may be indicative of a shift away from using context information that is selective to older adults.

Composites. The final comparison between NV and SV participants was made on the sensory and context composite variables, which are presented in Table 16. Examining the top panel of Table 16 reveals that the manipulation of signal quality had a strong effect on the extent to which participants responded on the basis of sensory information. Sensory responding increased in line with increases in signal quality, and the signal quality present in the NV condition seemed to serve well as an intermediate between the low and high SV conditions. In line with that finding, a 2 x (Training Group: SV, NV) x 2 (Age: Young, Older) x 2 (Signal Quality: Low, High) mixed-model repeated measures ANOVA on sensory composite scores revealed a significant training group x signal quality interaction, $F(1, 90) = 101.44$, $MSE = 2.33$, $p<.001$, $\eta_p^2 =.53$. Separate ANOVAs or the SV and NV training groups revealed that while NV participants showed no significant differences in sensory composite scores as a function of signal quality (as expected, because there was no actual manipulation in that condition), $F<1$, *ns*, there was a significant effect of signal quality for SV participants, $F(1, 44) = 6.29$, $MSE = .18$, $p<.05$, $\eta_p^2 =.13$. None of the above reported ANOVAs revealed significant main effects or interactions involving age.

While age groups did not significantly differ in terms of their sensory-based responding, the results presented in the bottom panel of Table 16 show that older adults were much more likely than young adults to respond on the basis of context. A 2 x (Training Group: SV, NV) x 2 (Age: Young, Older) x 2 (Signal Quality: Low, High) mixed-model repeated measures ANOVA on sensory composite scores revealed a strong effect of age, $F(1, 90) = 61.07$, $MSE = 4.44$, $p<.001$, $\eta_p^2 =.40$. A glance at the bottom panel of Table 16 also shows that SV had higher context composite scores in the low signal quality condition than the high signal quality condition. This led to a significant training group x signal quality interaction, $F(1, 90) = 10.30$, $MSE = .23$, $p<.01$, $\eta_p^2 =.10$, where context composite scores differed as a function of the signal quality manipulation in the SV condition, but not as a function of the arbitrary coding in the NV condition. The finding that context-based responding decreased as signal quality increased converges well with the calibration and monitoring results that showed that as sensory information was made more available in the SV training group, context effects were reduced.

Table 16. *Sensory-based and context-based responding composite for NV and SV participants during the training phase of Experiment 2.*

Signal Quality	SV - Low		NV		SV - High		Overall	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Sensory-Based								
Young	0.80	0.04	1.00	0.04	1.18	0.05	0.99	0.04
Older	0.71	0.05	0.97	0.05	1.17	0.04	0.95	0.05
Context-Based								
Young	0.91	0.05	0.89	0.03	0.82	0.05	0.87	0.04
Older	1.24	0.04	1.22	0.05	1.12	0.05	1.19	0.05

Summary

The pretest analyses presented at the beginning of this chapter found evidence for age differences in false hearing when the 0x procedure advocated at the end of Chapter 3 was used. Aside from having found that older adults were more confident in baseline hits than young adults, the 0x procedure succeeded in replicating past false hearing experiments. The pretest analyses also verified that random assignment of training groups was successful in ensuring a lack of training group differences, aside from one unexpected advantage in monitoring between CV and NV participants.

The regression analyses were conclusive across several measures in showing that variability training did not provide any advantage in reducing false hearing relative to a no variability control group. Though young and older adults both showed gains from pretest to posttest, there was not a significant difference that occurred as a function of training, excepting a small effect that suggested that CV training my be less effective for older adults than NV or SV training. On volunteered incongruent false alarms and on the context-based responding composite, evidence was obtained that suggested that older adults may show less positive training effects than young adults.

In addition to assessing the viability of variability training as a way of reducing false hearing, Experiment 2 was designed to test for age differences in sensitivity to the variability manipulations. Older adults were less sensitive to manipulations of context validity, as indicated by increased effects of context for older adults on monitoring when context validity was low. Young adults were also more likely to commit incongruent false alarms when context validity was high, indicating that they were sensitive to when context was more valid. This was in contrast to the findings regarding signal quality variability, that showed when context information was not available, older adults' confidence judgments were more sensitive to changes in signal quality than were young adults'. This finding may have important implications for future training, and is a rare example of an age-related increase in monitoring.

Chapter 6: General Discussion

The results of the current studies revealed some interesting new findings to be discussed regarding false hearing. Experiment 1 revealed that the priming observed by Rogers et al. (in prep) was solely because of the semantic relationship between the cue and target . Experiment 2 found that training incorporating variability of signal quality or context validity was not more effective at reducing false hearing compared to a practice-with-feedback control group. Age group differences in sensitivity of confident responses to the variability under which participants were trained were obtained in Experiment 2. Young adults had increased sensitivity relative to older adults to the context validity manipulation. Older adults showed increased sensitivity relative to young adults to the signal quality manipulation, but only when context information was not available as a basis for responding (i.e., baseline trials). Because work in false hearing is new and related only tangentially to a few literatures, the theoretical implications of the findings are limited. However, the practical applications of the work will be discussed further.

Is semantic priming responsible for false hearing?

The prior procedure used to elicit false hearing in older adults (Rogers et al., in prep) included a familiarization phase in which participants were to remember a set of cue-target pairs for a later memory test. All of the cue-target pairs were semantic associates and, so, subsequent priming could be the result of pre-existing semantic relationships. However, priming could also have originated from memory for the cue-target pairings presented during training. That is, the episodic and semantic components of the priming manipulation were confounded. The results of Experiment 1 showed that increasing repetitions did not serve to increase false hearing as measured by either accuracy or subjective experience. In the 0x condition, where context effects could only be driven by semantic priming, the pattern of older adults being more reliant upon contextual information than young adults was still obtained: older adults had more congruent hits and incongruent false alarms than did young adults, greater confidence in their congruent hits and incongruent false alarms, and had poorer monitoring and calibration than young adults on incongruent trials.

In Experiment 1, participants were least confident in their incongruent false alarms when the cue had been presented five times during the training phase. In fact, the older adults were most likely to falsely hear a word favored by context when it had not been presented during the training phase. This finding strongly suggests that false hearing resulted from semantic priming. Repeated presentations during training did not increase context-based responding but, rather, may have swayed participants toward requiring more sensory information to endorse the response favored by context with high confidence. When the word favored by context was presented in noise, repeated presentations were associated with higher confidence in congruent hits. In contrast, when the favored word was not presented, repeated presentations were associated with lower confidence in incongruent false alarms. This finding suggests that there may have been an episodic effect on priming, but that it operated in a manner that encouraged sensory-based responding and thereby reduced false hearing.

Using a method similar to that of Experiment 1, one could test if a repetition-based priming manipulation could increase context-based responding beyond semantic priming. To do this, the targets and alternates during the perceptual test phase should be both semantically related to the cue. Recall that in Experiment 1, the alternate response was not semantically related to the cue (e.g., ROW- GOAT). Given a training phase where repeated presentations would be manipulated (e.g., 0x, 3x, and 5x), would repetitions increase false hearing if trained on "drunk-rum" and then "drunk-bum" is presented as an incongruent trial? In a 0x situation, the probability of choosing "rum" or "bum" may be very similar (to the extent that the forward-acting associative strengths from the cue are the same), but if "drunk-rum" is repeatedly presented, "rum" may be more likely to be invalidly reported, and potentially with high confidence. If older adults were more likely to show false hearing in this situation, it would suggest that repetition priming increases a familiarity-derived basis for false hearing.

Implications for training reductions in false hearing

The findings in Experiment 2 regarding the efficacy of variability training as a method for reducing false hearing were discouraging in that the CV and SV training groups did not show greater positive effects of training than NV participants. However, it is not the case that training in Experiment 2 failed to reduce false hearing. Rather, the training groups did not differ from one another—aside from older adults in the CV group who showed slightly less benefits of training than all other participants. That said, false hearing was not eliminated altogether with feedback. The patterns of contextual reliance shown by older adults in these listening tasks are robust, resistant to change, and clearly not a result of older adults misunderstanding the task instructions.

A likely possibility is that the amount of training administered in Experiment 2 was not sufficient to show different effects for the three training conditions. Experiment 2 comprised a single two-hour session for each participant, including pretest and posttest. This is a very small training study when compared to the other training studies reviewed in the introduction (e.g., Baron & Surdy, 1990; Jennings & Jacoby, 2003; Aizpurua & Koutstaal, 2010) that included several training sessions that spanned over weeks, or in some cases months. It also should be noted that one of the motivating studies for this thesis (Sommers & Barcroft, 2005) demonstrated an advantage for variability in speakers when training Spanish vocabulary learning after a short 1-hour long experiment. However, this exception suggests that the kind of variability produced by varying the number of speakers in Sommers and Barcroft (2005) may be very different from the variability in context validity and signal quality reported here. In this study, all materials were generated from a single speaker. These studies, as well as those done by the Advanced Cognitive Training for Independent and Vital Elderly group (ACTIVE, e.g., Ball et. al., 2002), were not concerned with audition or the subjective experience of audition, but were concerned primarily with memory training.

The last possibility to discuss in terms of why the variability manipulation may not have produced positive training effects is that the procedure did not include a training-to-criterion or shaping procedure that would have taken into account each participant's performance, and then tailored the amount of variability that each participant received. As an example, several successful training studies (e.g., Jennings & Jacoby, 2003; Mahncke et. al., 2006) used shaping procedures that forced participants to remain at a particular difficulty level until a certain criterion level of performance was reached, at which point the level of difficulty was incremented. Shaping procedures are specifically designed to change behavior, but may also inform subjective experience to be sensitive to that change. Shaping reductions in false hearing may encourage sensitivity to change for greater generalization to new situations. Improved monitoring and calibration may lead to greater broad transfer to new situations (Weed, Ryan, & Day, 1990), increasing the value of training.

Age differences in sensitivity to variability

Though Experiment 2 failed to demonstrate greater efficacy of variability training in reducing false hearing relative to a no variability control group, Experiment 2 did reveal some new findings regarding age differences in sensitivity to the variability that was present during training. The patterns of results were different for variability in context validity than for signal variability. I will discuss each and their implications in turn.

For the context validity manipulation, comparing CV and NV participants, the results showed that older adults' monitoring was much poorer than young adults' when context validity was low. Young adults also were more likely to commit an incongruent false alarm when context validity was high, showing they shifted toward responding on the basis of context when it was more valid. This finding of an age group difference was surprising given that both groups were exposed to feedback after every trial, which should have indicated to both groups equally how often context was valid. Two explanations could account for why older participants were not as sensitive to the validity manipulations as young adults. The first is that participants were only informed about the variability in context validity indirectly through feedback. Keeping track of how often the contextual cue was valid or invalid in addition to identifying the word in noise could have created a dual-task situation. Such situations favor young adults (Verhaeghen, et. al., 2003), and may have limited older adults' awareness of the variability. If the manipulation of validity had been presented in a more salient manner, then potentially there would not have been an age group difference in sensitivity to variability. For example, the context validity manipulation could have been presented through two different speakers. During the identification task one speaker could have been used for the high context validity condition, and another speaker for the low context validity condition. Participants then could have been warned to trust one speaker, and not another. Testing in this way would improve salience of the manipulation and free participants from having to calculate the validity of context online.

A second possibility for why older adults may not have been as sensitive to the context validity manipulation as young adults could have been that older adults were captured by the response favored by context, and then assigned high confidence without pausing to consider whether or not the contextual information was valid. Jacoby et al. (2005) found in memory that older adults were more likely to be captured by an invalidly primed response than young adults, leading to dramatic false remembering. Rogers et al. (in prep) suggested that false hearing could operate in a similar manner. If false hearing does work in this way, then training aimed toward constraining listening to a particular source of information may reduce the likelihood of being captured by another invalid and irrelevant source of information.

Turning to signal validity, in the SV condition of Experiment 2, both age groups had equal increases in congruent hit rate and decreases in incongruent false alarm rate when signal quality was improved from low to high. Participants' confidence in these responses did not significantly change when signal quality was improved, potentially because when responses were favored by context, sensory information need not have been a basis for confidence. When analyses were not conditionalized by response type, overall confidence in the congruent condition increased to the same extent as did accuracy, but for incongruent trials, participants' increase in confidence greatly underestimated the increase in accuracy. Age groups did not differ in this lack of sensitivity in incongruent trials, and showed similar improvements in calibration bias from low to high signal quality.

The lack of sensitivity to increases in signal quality could be indicative of why subjective customer satisfaction with hearing amplification has not improved for 30 years, despite the radical innovations in hearing aid technology (Kochkin, 2003). If hearing aid users also base their confidence in hearing on contextual information, then increases in signal quality caused by the hearing aid may only be noticed in the rare situation when context information is not available. This may lead to the impression that the hearing aids do not work, which may influence a person's decision to not use their hearing aid.

Experiment 2 also revealed that when context was not available as a basis for responding (e.g., baseline trials), older adults' confidence ratings were more sensitive than young adults to variability in signal quality. This pattern was obtained when baseline confidence ratings were conditionalized for hits only and when they were not conditionalized. This finding could suggest than in everyday listening situations older adults are more aware of the quality of listening conditions, but only when context is not available. Thus, older adults may be more likely to sort preference in restaurants on the basis of noise than young adults. Older adults may also be more likely than young adults to try to establish context early in conversation, so as to reduce demands on sensory-based responding. Potentially, older adults' subjective experience of hearing could be greatly improved with one high fidelity topic sentence at the beginning of a conversation to establish context. For example, older adults could be more likely than young adults to initiate conversations with main idea statements, such as "Did you watch the football game yesterday?" instead of "Manning had a great game yesterday, didn't he?" Once context is established however, the risk of false hearing still looms.

Conclusion

The present research described two variability-based training procedures designed to reduce false hearing in older adults that were tested against a no-variability control treatment. Though the results showed that neither of the variability-based training groups performed better than the control group, they did reveal novel effects regarding age differences in sensitivity to variability in signal quality and context validity. Before the training study was conducted, a preliminary experiment showed that the procedures used by Rogers et al. (in prep) to elicit false hearing were capitalizing on the effects of semantic priming. These findings should be important for future training studies aimed at reducing false hearing in the elderly, which was shown to be robust even after feedback.

Older adults' greater reliance on context is certainly adaptive in many situations. In most naturalistic listening situations, context is valid. This work does not advocate the idea that older adults should abandon use of context when listening in daily life. Rather, this work serves to highlight that there are multiple qualitatively different bases for hearing, and that older adults' use of context comes at a cost to perceptual discriminability. In situations where context is invalid, this can lead to false hearing. Just as the work in false memory does not advocate abandoning use of memory, this work in false hearing does not advocate abandoning use of context. Rather, the target for training is careful listening and being flexible across the different bases for hearing.

References

Aizpurua, A., & Koutstaal, W. (2010). Aging and flexible remembering: Contributions of conceptual span, fluid intelligence, and frontal functioning. *Psychology and Aging*, *25*(1), 193-207. doi:10.1037/a0018198

American Speech-Language-Hearing Association. (1988). *Determining Threshold Level for Speech* [Guidelines].

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., & Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA: Journal of the American Medical Association*, *288*(18), 2271-2281. doi:10.1001/jama.288.18.2271

Barcroft, J., & Sommers, M. S. (2005). Effects of Acoustic Variability on Second Language Vocabulary Learning. *Studies in Second Language Acquisition, 27*(3), 387-414. doi:10.1017/S0272263105050175

Baron, A., & Surdy, T. M. (1990). Recognition memory in older adults: Adjustment to changing contingencies. *Journal of the Experimental Analysis of Behavior*, *54*(3), 201-212. doi:10.1901/jeab.1990.54-201

Benjamin, A. S. & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), Implicit memory and metacognition: The 27th Carnegie Symposium on Cognition (pp. 309-338). Hillsdale, NJ: Erlbaum.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*(2), 123-152. doi:10.1037/h0043805

Cattell, R. B., & Cattell, A. K. S. (1960). *Handbook for the individual or group Culture Fair Intelligence Test*. Champaign, IL: IPAT.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Oxford England: Lawrence Erlbaum.

Dimitrov, D. M., & Rumrill, P. D., Jr. (2003). Pretest-posttest designs and measurement of change. *Work: Journal of Prevention, Assessment & Rehabilitation, 20*(2), 159-165.

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*(1), 538-546. doi:10.1121/1.428322

Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition, 27*(2), 328-338.

Goode, M.K., Geraci, L., & Roediger, H.L. (2008). Superiority of variable to repeated practice in transfer on anagram solution. *Psychonomic Bulletin & Review, 15*, 662-666.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford England: Wiley.

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: Advances in* *research and theory, Vol. 22*. (pp. 193-225). San Diego, CA US: 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*(1), 122-134. doi:10.1037/0882-7974.14.1.122

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica, 26*(2), 107-129.

Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin, 82*(4), 511-518. doi:10.1037/h0076767

Hutchinson, K. M. (1989). Influence of sentence context on speech perception in young and older adults. *Journals of Gerontology, 44*(2), P36-PP44.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 3-22. doi:10.1037/0278-7393.25.1.3

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 240-247. doi:10.1037/0278-7393.14.2.240

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, Subjective Experience, and Cognitive Control: Dramatic False Remembering by Older Adults. *Journal of Experimental Psychology: General, 134*(2), 131-148. doi:10.1037/0096-3445.134.2.131

Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subject reports of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 686-700. doi:10.1037/0278-7393.27.3.686

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. *Varieties of memory and consciousness: Essays in honour of Endel Tulving.* (pp. 391-422). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Jennings, E. (1988). Models for pretest-posttest data: Repeated measures ANOVA revisited. *Journal of Educational Statistics, 13*(3), 273-280. doi:10.2307/1164655

Jennings, J. M., & Jacoby, L. L. (2003). Improving memory in older adults: Training recollection. *Neuropsychological Rehabilitation, 13*(4), 417.

Jennings, J. M., Webster, L. M., Kleykamp, B. A., & Dagenbach, D. (2005). Recollection Training and Transfer Effects in Older Adults: Successful Use of a Repetition-Lag Procedure. *Aging, Neuropsychology & Cognition, 12*(3), 278-298. doi:10.1080/138255890968312

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*(1), 1-24. doi:10.1006/jmla.1993.1001

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory & Language, 48*(4), 704. doi:10.1016/S0749-596X(02)00504-1

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills, 46*(2), 395-401.

Kochkin, S. (2003). MarkeTrak VI: On the Issue of Value: Hearing Aid Benefit, Price, Satisfaction, and Repurchase Rates. *The Hearing Review, 10*(2), 12.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*(3), 297-315. doi:10.1037/0096-3445.123.3.297

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517. doi:10.1037/0033-295X.103.3.490

Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review, 13*(1), 84-91.

Kramer, A.F., Larish, J.F., & Strayer, D.L. (1999). Training for executive control in dual-task settings: A comparison of young and old adults. *Journal of Experimental Psychology: Applied. 1*, 50-76.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning & Verbal Behavior, 13*(5), 585-589. doi:10.1016/S0022-5371(74)80011-3

Lyle, K. B., Bloise, S. M., & Johnson, M. K. (2006). Age-related binding deficits and the content of false memories. *Psychology and Aging, 21*(1), 86-95. doi:10.1037/0882-7974.21.1.86

Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., Joyce, N. M., Boniske, T., Atkins, S. M., & Merzenich, M. M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences of the United States of America, 103*(33), 12523-12528. doi:10.1073/pnas.0605194103

Miller, G. M., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*(1), 40-48. doi:10.1037/0021-843X.110.1.40

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1170-1187.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402-407.

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*(6), 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*(6), 838.

Pichora-Fuller, M. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology, 47*, 72-82. 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*(1), 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*(2), 275-280. doi:10.1037/h0027768

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814. doi:10.1037/0278-7393.21.4.803

Rogers, C.S., Jacoby, L.L., & Sommers, M.S. (in preparation). A Metacognitive Analysis of Age Differences in False Hearing.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49*(4), 304-313. doi:10.1037/0003-066X.49.4.304

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217. doi:10.1111/j.1467-9280.1992.tb00029.x

Shipley, W. C. (1986). *Shipley Institute of Living Scale*.

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*(2), 333-341. doi:10.1037/0882-7974.11.2.333

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), 458-472. doi:10.1037/0882-7974.14.3.458

Sweetow, R., & Palmer, C. V. (2005). Efficacy of Individual Auditory Training in Adults: A Systematic Review of the Evidence. *Journal of the American Academy of Audiology, 16*(7), 494-504.

Thapar, A., & Westerman, D. L. (2009). Aging and fluency-based illusions in recognition memory. *Psychology and Aging, 24*(3), 595-603. doi:10.1037/a0016575

Tranter, L. J., & Koutstaal, W. (2008). Age and Flexible Thinking: An Experimental Demonstration of the Beneficial Effects of Increased Cognitively Stimulating Activity on Fluid Intelligence in Healthy Older Adults. *Aging, Neuropsychology & Cognition, 15*(2), 184-207. doi:10.1080/13825580701322163

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology, 59*(9), 920-925. doi:10.1016/j.jclinepi.2006.02.007

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A meta-analysis. *Psychology and Aging, 18*(3), 443-460. doi:10.1037/0882-7974.18.3.443

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised manual*. New York, NY: Psychological Corporation.

Weed, K., Ryan, E. B., & Day, J. (1990). Metamemory and attributions as mediators of strategy use and recall. *Journal of Educational Psychology, 82*(4), 849-855. doi:10.1037/0022-0663.82.4.849

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1*, 59-85.

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*(3), 144-148. doi:10.1111/j.0963-7214.2005.00356.x

Appendix: List of Stimuli

The first set of triples corresponds to the word triples used for the congruent and incongruent conditions in Experiments 1 and 2. The two-word pairs were used for baseline trials.

cue	target	alternate
front	back	bat
good	bad	bag
bounce	ball	doll
tub	bath	bass
laser	beam	deem
grizzly	bear	dare
wager	bet	get
small	big	bid
pedal	bike	bite
row	boat	goat
insect	bug	bud
hobo	bum	gum
taxi	cab	tab
icing	cake	cape
auto	car	par
dog	cat	cap
effect	cause	pause
cavern	cave	pave
inexpensive	cheap	cheek
miner	coal	pole
jacket	coat	cope
corn	cob	cod
morse	code	toad
chef	cook	took
sofa	couch	pouch
saucer	cup	cut
calendar	date	bait
dusk	dawn	gone
alive	dead	bed
life	death	deaf
shallow	deep	beep
doe	deer	gear
knob	door	boar
up	down	gown
quack	duck	buck
flunk	fail	sail
handbag	purse	curse
skinny	fat	sat
touch	feel	seal
mist	fog	hog

against	for	soar
fuel	gas	gaffe
gander	goose	goof
pistol	gun	bun
brush	hair	fair
corridor	hall	fall
love	hate	fate
hat	head	fed
there	here	fear
steep	hill	fill
house	home	foam
hula	hoop	hoot
cold	hot	hop
embrace	hug	hub
career	job	jog
retain	keep	peep
yarn	knit	nip
tardy	late	lake
giggle	laugh	lass
follow	lead	league
arm	leg	led
lamp	light	like
cabin	log	lob
win	lose	luge
atlas	map	mat
calculus	math	mass
tongue	mouth	mouse
day	night	knife
hurt	pain	cane
skillet	pan	tan
pencil	pen	ten
choose	pick	kick
chlorine	pool	tool
rich	poor	tore
rodent	rat	rap
coral	reef	wreath
left	right	ripe
street	road	robe
stone	rock	rot
knot	rope	rote
thorn	rose	rove
happy	sad	sag

different	same	fame
buy	sell	fell
shepard	sheep	sheet
daughter	son	fun
total	sum	thumb
wag	tail	pail
short	tall	call
instruct	teach	peach
rip	tear	pair
thick	thin	fin
object	thing	sing
feet	toes	pose
bottom	top	taught
run	walk	watt
strong	weak	weep
scale	weight	wake
black	white	wipe
half	whole	foal
better	worse	worth
socks	shoes	sues
smooth	rough	rush
spider	web	wed
carpet	rug	rub
blaze	fire	sire
blackboard	chalk	chop
murder	kill	pill
butcher	meat	meek
hymn	song	thong
freckle	face	faith
ill	sick	sip
chime	bell	dell
rhythm	beat	beak
shove	push	puss
rescue	save	shave
lather	soap	hope
monopoly	game	dame

Baseline pairs

cue	target
brace	dot
leash	mock
cement	kite
jaw	pass
strengthen	bead
friction	bean
absent	bowl
rigid	mud
blend	height
normal	bill
sorrow	thud
bigot	bide
pillow	came
acrobat	chick
dreamed	cage
hound	lop
crawl	catch
calcium	calm
vapor	leak
stable	kit
brief	lip
hotel	rob
dance	veal
safe	cone
summer	coach
brought	moat
band	keys
range	dull
mode	boom
finally	suck
creature	bash
circle	dig
grow	coil
pull	hide
oil	keen
carry	keel
explain	caught
solution	fine
chance	foil
party	bile
result	folk
dark	leaf
clear	chafe
water	but
mind	faze
business	five
until	foul
tell	hone
year	heard
well	bang
first	heed
should	suit
set	shop
end	lack
service	nab
found	curl

area	hit
note	make
become	roof
early	rid
pamper	lad
northward	neck
allure	take
halter	bathe
heartless	loop
hedge	raise
clap	gaze
washer	muck
sausage	keg
drip	pile
revert	pack
bamboo	seam
popcorn	pun
spruce	tyke
tremor	gait
flurry	rise
maim	chat
fume	load
smother	rake
razor	sack
stole	moss
tender	seat
colony	fit
stern	cite
pulse	pup
assertion	fang
powder	cub
prayer	puck
grace	kin
dice	peer
weird	pap
costume	fought
bloom	case
slick	tip
squad	gnat
pepper	might
cotton	shake
nurse	wick
extend	foot
motor	sash
shoulder	sock
fuchsia	shone
spot	heat
lenient	peak

procession	hack
fiend	sake
suburb	said
refute	serve
rash	mess
rim	shut
slug	wish
refund	hutch
theft	hung
comedy	shade
infant	sear
trail	chic
upset	hook