The Effects of Context and Feedback on Age Differences in Spoken Word Recognition

Elizabeth A. L. Stine-Morrow, Lisa M. Soederberg Miller, and John A. Nevin

We investigated the hypothesis that age differences in speech discrimination would be reduced by enhancing the distinctiveness of the speech processing event in terms of both the context of encoding and the response outcome. Younger and older adults performed an auditory lexical decision task in which the degree of semantic constraint (context) and type of feedback were manipulated. Main effects of age indicated that older adults generally showed lower discriminability (D) and greater bias (B) toward reporting signals to be words. Consistent with the environmental support hypothesis, older adults were differentially facilitated in discriminability by feedback, but only when semantic context was provided. Also, for both younger and older adults, feedback and context each had the effect of reducing bias and facilitating the speed of rejecting nonwords. Contrary to one suggestion in the literature that aging brings an insensitivity to environmental contingency, older adults were at least as capable as the young in taking advantage of feedback to normalize the speech signal so as to increase discriminability and decrease bias.

The first time I went over she called me “my dear,” . . . only she said, “My deah.” For the longest time I couldn’t figure out what she was saying. I thought she was talking about an idea she had, but it never made sense in the sentence. She said, “My deah, why don’t you stay and listen? This book is one of my favorites.”

—Jane Hamilton, The Book of Ruth

In processing spoken language, the listener constructs meaning from a continuous acoustic stream. Even under the best of circumstances, this signal is only fleetingly available and varies in acoustic form between speakers (Mullennix, Pisoni, & Martin, 1989). Furthermore, individual speech sounds (phonemes) are coarticulated, such that these units overlap in their production across and within words, resulting in a signal whose acoustic properties vary as a function of the surrounding speech context. Given such an inherently muddy and variable signal, a primary problem facing researchers is how individuals recognize and understand words. The question becomes particularly acute in the context of gerontology in which comprehension of an elusive and variable signal is further hindered by the pervasive declines in sensory and cognitive systems that occur as a normal part of aging (Working Group on Speech Understanding and Aging, 1988).

A survey of the recent literature shows that this issue has by no means been neglected (see Wingfield, 1999, and Wingfield & Stine-Morrow, in press, for reviews). Much of this work has centered on the nature of the speech conditions that can exacerbate or ameliorate age differences in speech understanding. One factor that has often been shown to mitigate age deficits in speech processing is a highly constraining semantic context (Cohen & Faulkner, 1983; Pichora-Fuller, Schneider, & Daneman, 1995; Wingfield, Aberdeen, & Stine, 1991; Wingfield, Poon, Lombardi, & Lowe, 1985). This age by context interaction is not always found, however (Craik, 1968; Holtzman, Familitant, Deptula, & Hoyer, 1986), and though it has been suggested that task and memory demands may moderate the existence of this interaction (Holtzman et al., 1986; Wingfield, Alexander, & Cavigelli, 1994), the nature of these moderators remains to be specified fully.

There is considerable debate in the cognitive literature with respect to how context influences word recognition. Although most models of word recognition share the assumptions that (a) multiple candidates from the mental lexicon are activated early on as the acoustic input unfolds in time, and (b) the acoustic input is matched against a stored representation of the phonological structure of the word, it is unclear how context operates on these processes. For example, it is uncertain whether context only comes into play relatively late after data-driven processes have already activated an array of lexical candidates (as argued in “modular” models) or early in conjunction with data-driven processes (as argued in “interactive” models; for a review, see Lively, Pisoni, & Goldinger, 1994). Regardless of the time course, however, the explanation in the cognitive aging literature for elders’ greater reliance on context (when it is found) is that the activation of lexical candidates by the context enables older adults to “fill in” the gaps created in the auditory input by an impoverished auditory system (cf. Holtzman et al., 1986; Pichora-Fuller et al., 1995; Wingfield et al., 1991).

Another factor of potential importance—one that has received very little attention in the cognitive aging literature (as well as in the more general cognitive literature)—is the effect of feedback, or outcome, on speech recognition. That is, in the ordinary course of language understanding, we access meanings of particular words; sometimes we are right and sometimes we realize we are not. Consider two possible interactions between two men, both moderately hard of hearing, at the beginning of a social conversation. The host says either, “Will you take a drink?” (let’s call this event S1) or “Will it rain, ya think?” (S2). The guest replies either, “I’d like a li’l tea” (B1) or “It’s likely—we will see” (B2). The host then responds either, “I have English Breakfast” (R1) or “I have a spare umbrella” (R2). Obviously, this interaction is successful only if the guest’s reply fits with the host’s initial question, and if the host’s response fits with the guest’s reply.
Confusion can arise at either point in the exchange, depending on how well the individual words are understood within the available context. For example, the production of B2 following on S1 would be likely to result in a puzzled expression on the part of the host (and certainly neither R1 nor R2). The host’s reaction would, in turn, cause the guest to realize that there must have been some problem with his on-line understanding of the initial query. On reflection, he may even be able to reconstruct that he was asked about refreshments rather than the weather. Sometimes such disambiguation is strategic as in the preceding example in which participants become conscious of a communication failure, but sometimes it is not, as in the case in which we learn the nuances of a novel speaker’s productions (Mullenix et al., 1989). To put the matter more generally, effective interaction between two speakers depends on the reply being appropriate to the previous utterance. Feedback (R1 or R2) helps us achieve perceptual constancy in the face of ordinary acoustic variability.

Our goal in the present study was to consider the effects of both context and outcome on age differences in speech recognition. As exemplars of environmental support (Craik & Jennings, 1992), a highly constraining semantic context at encoding and a distinctive outcome (which would provide the listener feedback about the correctness of lexical access as well as the perceptual constancy of the speech signal) could be especially important in moderating age differences in speech recognition. These issues become interesting in the context of arguments in the literature that suggest that older adults may take either greater (Craik & Jennings, 1992) or lesser (Baron & Surdy, 1990) advantage of such factors.

**Lexical Decisions as Concurrent Discriminated Operants**

In order to capture the effects of outcome in the speech processing event in the laboratory, we used an auditory lexical decision task and conceptualized performance in terms of the “concurrent discriminated operant” (Davison & Nevin, in press; Skinner, 1969). In such a framework, behavior and its context is described as a three-term contingency: the stimulus (S; in this case, the acoustic event); the behavior, or response (B; in everyday discourse processing, lexical access; in our laboratory analogue, an overt “word”/“nonword” judgment); and the reinforcer (R; in everyday discourse processing, an “appropriate” response from a listener, or perhaps, the construction of coherent meaning; in our laboratory analogue, explicit feedback with regard to whether the lexical decision was correct). Behavioral plasticity is possible to the extent that organisms discriminate among various S-B relations (e.g., responding differently and appropriately to an offer of a drink vs small talk about the weather) and among various B-R relations (e.g., completing a successful communicative interaction).

Applying this to our current problem, we note that speech understanding problems among elderly adults could arise because of a failure to discriminate among either S-B relations or B-R relations. The natural assumption, the one ordinarily made in the literature, is that age-related difficulties arise because of deficits in the former system, but B-R relations may also be a contributing factor. For example, we note anecdotally the tendency to respond to a nonsensical utterance produced by an elderly adult who is perceived to be cognitively failing as though it were ordinary (“well yes, if you say so”); it is tempting to attribute continuing cognitive decline to weak or indiscernible B-R relations. In fact, there is ample empirical evidence among elderly long-term care residents that dependency can arise from a social environment that fails to reinforce independent behaviors (M. Baltes, Kindermann, Reiszeben, & Schmid, 1987; M. Baltes, Neumann, & Zank, 1994; P. Baltes & M. Baltes, 1990). In other words, our ability to successfully comprehend and communicate (which is an important feature of an independent lifestyle) depends on the availability of veridical feedback on the accuracy of our understanding; this could arguably become more important in the face of age-related declines in cognition. So, in experimental assessments of speech understanding, to the extent that B-R relations are neglected—or controlled so as to be highly confusable—deficits in the ability to discriminate among S-B relations might be exaggerated.

This principle can be understood in terms of the discriminative operant model of Davison and Nevin (in press), which was inspired in part by signal-detection theory (for review, see Luce, 1963). In psychophysics, the major achievement of signal-detection theory was the isolation of a single parameter, $d’$, that varied systematically with signal-to-noise ratio but remained constant, for individual subjects, both with respect to variables that affected response bias and across measurement paradigms (Swets, Tanner, & Birdsall, 1961). Accordingly, $d’$ was construed as a fundamental measure of signal discriminability. Subsequently, signal-detection analyses were applied quite broadly (for review, see Hutchinson, 1981), and the importance of separating stimulus discrimination and response bias has been recognized in a number of areas including recognition memory (e.g., Murdock, 1982), social psychology (e.g., Martin & Rovira, 1981), and animal discrimination learning (e.g., Heinemann & Chase, 1975). Most recently, Alsop (1998) has argued that “signal-detection performance is the product of a variety of discriminations involving the sample stimuli, the response alternatives, and the feedback or outcomes for these choices” (p. 249) and that all these factors must be taken into account in experiments that compare the performance of independent groups of subjects. The Davison-Nevin model formalizes and quantifies the roles of these factors as they affect both stimulus discrimination and response bias. Here, we apply this approach to the analysis of age differences in word recognition.

Discrimination between S1 and S2 (in this case, words and nonwords) can be separated out from a response bias toward B1 or B2 (in this case, “Yes, the acoustic pattern is a word” or “No, it is not”), with discrimination measured as:

$$D = \{[p(B1|S1)/p(B2|S1)] \cdot [p(B2|S2)/p(B1|S2)]\}^{0.5}$$

and bias measured as:

$$B = \{[p(B1|S1)/p(B2|S1)] \cdot [p(B1|S2)/p(B2|S2)]\}^{0.5}$$

The model assumes that when S1 is presented and the subject responds B1 (correct), the tendency to repeat B1 on the next S1 presentation is strengthened by its immediate consequence R1. Thus, discrimination between confusable signals (S1 vs S2) is shaped by its consequences. However, responses and their consequences (i.e., reinforcers or feedback) may themselves be confusable, thereby decreasing discrimination (D). For example, Nevin, Cat, and Alsop (1993) demonstrated that D could be enhanced similarly by increasing the difference between the stimuli (S1 and S2).
or the difference between the topography of the behaviors (B1 and B2; e.g., a response with a short latency for "yes" or with a long latency for "no"). Another way to enhance discrimination performance (as we will elaborate on in a moment) is to arrange qualitatively different outcomes for the two types of correct responses (SB11 and SB22; e.g., Trapold, 1970; see Peterson, 1984, for a review). Taken together, these results suggest that discrimination depends on the distinctiveness of the entire three-term S-B-R contingency which defines the discriminated operand.

In a typical auditory lexical decision task, there are no contingencies on behavior at all because the assumption is that it is the S-B relation that is the primary determinant of performance. The lack of feedback ("a no feedback [NF] condition"); then, creates a situation in which the distinctiveness of B-R relations is minimal. We could modify this, of course, to produce a simple feedback situation in which hits (achieving lexical access when in fact there has been a word, and indicating "word") and correct rejections (not accessing a lexical entry when there has not been a word, and indicating "nonword") result in one kind of outcome (e.g., a "correct" signal) while false alarms and misses result in another kind of outcome (e.g., an "incorrect" signal). We will refer to this as a same feedback (SF) condition. Because the S-B relations (SB11 and SB22) occasion one kind of outcome, whereas the other two (SB12 and SB21) occur another, the distinctiveness of B-R relations is enhanced relative to the NF situation, but it is still not optimum: there is still some confusability between SB11 and SB22. The alternative would be to reinforce differentially the two kinds of correct responses (i.e., one kind of feedback for SB11 and a different kind of feedback for SB22) so as to further enhance the discriminability of the B-R relation, a paradigm referred to as differential feedback (DF). In fact, animal work has demonstrated that such differential feedback produces faster learning, a phenomenon termed the differential outcome effect (Trapold, 1970). Davison and Nevin (in press) argue that this increase in the distinctiveness of B-R relations increases the distinctiveness of the entire discriminated operand. Thus, increasing the difference between outcomes may be functionally equivalent to increasing the differences between the stimuli themselves, a principle of potential importance to gerontology because aging appears to bring reliable declines in sensory discrimination (Baltes & Lindenberger, 1997).

Varying the terms of the discriminated operand can also affect bias. When there is no external feedback (NF), bias may arise if the value of self-generated feedback is greater for one of the correct responses, SB11 or SB22. For example, it may be easier or more satisfying to recognize that one has correctly identified a word than a nonword, leading to a general bias toward "word" responses. When explicit feedback is arranged equally often for the two correct responses, self-generated feedback would be relatively less distinctive or valued, and bias should conform approximately to the veridical relative frequency of words and nonwords. Thus, if participants exhibit an inherent bias, it should be reduced by feedback, and the feedback effect should be larger as the distinctiveness of B-R relations is increased (e.g., SF vs DF).

Implications of S-B and B-R Distinctiveness for Cognitive Aging

Even though signal detection has been used for some time to examine discriminability and bias as separable contributors to age differences in performance (e.g., Perry, 1993; Poon & Fozard, 1980; Potash & Jones, 1977), the contributions of B-R relations have been neglected. In fact, the discriminated operand framework meshes well with some conceptualizations of cognitive aging. For example, some theorists (Craik & Jennings, 1992; Rabinowitz & Ackerman, 1982) have argued that aging brings a decrease in the ability to encode events distinctively. To the extent that the distinctiveness hypothesis is correct, this principle would be expected to extend to the encoding of the entire operand—to B-R relations as well as to S-B relations. As we have already noted, if this is the case, then experiments that assess age differences in discrimination with either no outcomes or nondifferential outcomes may exacerbate age differences by making the entire event less distinctive. Thus, age differences in sensitivity to S-B relations may be overestimated.

The idea, then, is that feedback might be especially helpful to elderly adults in a discrimination task because it would augment the distinctiveness of B-R relations, thereby increasing environmental support (Craik & Jennings, 1992).

The effects of the distinctiveness of outcome on age differences is interesting from another standpoint as well. There has been some suggestion in the literature that aging brings a decreased sensitivity to outcome (Baron & Surdy, 1990). To the extent that older adults are less sensitive to consequences, learning itself would, of course, be compromised (Freund, 1996; Kausler, 1991). Baron and Surdy (1990) examined the performance of four older and four younger men in a continuous recognition experiment in which the payoffs for “new” (nonrecognition) and “old” (recognition) responses were manipulated. The payoff schedule was either balanced between recognition and nonrecognition, structured with a recognition-bias (heavy rewards for hits and heavy penalties for misses), or structured with a nonrecognition-bias (heavy rewards for correct rejections and heavy penalties for false alarms). Although the bias among the younger men was strongly influenced by the payoff schedule, the older men showed a much weaker effect, leading these investigators to suggest that aging brings a “reduced sensitivity to changed contingencies” because of the “more extensive history” of the older adult (p. 208). If the Baron and Surdy conceptualization is correct, we would expect older adults to show less benefit from same and differential feedback than would young adults.

To summarize our basic argument, we investigated the effects of sentence context and feedback in an auditory lexical decision task. Based on Craik's notion that aging brings increased reliance on environmental support to overcome a decrease in the ability to self-initiate processing, we hypothesized that the presence of context and feedback would show larger beneficial effects among older adults than among young adults. In particular, in examining the effects of feedback, we were interested in eliciting the environmental support idea against the antithesis proposed by Baron and Surdy that aging brings a reduced sensitivity to behavioral consequences.

Methods

Participants

Participants were 61 younger (M = 22.4 years; SD = 4.2) and 60 older (M = 70.0 years; SD = 4.9) adults. Younger adults were primarily, and elderly adults were exclusively, recruited from the community. Though some younger partici-
pants were college students, none received college credit or bonus points; all participants received compensation that was completely contingent on their task performance. Older adults were higher in WAIS vocabulary ($M_{OA} = 60.4; SD_{OA} = 5.2$) than were the younger adults ($M_{YA} = 54.6; SD_{YA} = 6.8$), $F(1,106) = 25.86, p < .0001$. They were also on average higher in educational level ($M_{OA} = 16.2; SD_{OA} = .3$) than were the younger adults ($M_{YA} = 14.3; SD_{YA} = .2$), $F(1,99) = 26.09, p < .0001$. The younger and older groups were not significantly different in either forward or backward digit spans, however, $F < 1$ for both.

The first half of the participants were tested in an experiment without contextual constraints (No-Context condition; NC), and the second half were tested with contextual constraints (Context condition, C). The results of these two experiments are combined into a single analysis. One third of the participants from each age-context group (Y-C, Y-NC, O-C, O-NO) were randomly assigned to one of three feedback conditions, No Feedback, Same Feedback, and Differential Feedback, resulting in 10 participants in each group, except for the Young-Context-Differential group, which had 11. A series of Age X Context X Feedback Condition ANOVAs showed that neither context nor feedback groups differed in WAIS vocabulary, or forward or backward digit span.

An audiometric analysis was not conducted on these participants. None of them used a hearing aid, and all reported their hearing to be “excellent” or “good” on a 5-point scale. As we will show, however, when words were presented without feedback or context, there were large age differences in the ability to discriminate between words and nonwords, so we can be fairly sure that the elderly adults had reduced hearing sensitivity, which is typical of this population (Working Group on Speech Understanding and Aging, 1988). Because participants were randomly assigned to groups, we assume that auditory deficits were similarly distributed in all experimental conditions.

Materials

Stimuli were words and nonwords that were embedded in one of two levels of white noise and were presented either with or without sentential context. This was accomplished in the following way.

Stimulus items were constructed out of an original list of 400 one- or two-syllable words, each of which was changed into a corresponding nonword by replacing one phoneme with another phoneme from the same phoneme class (e.g., a labial stop like /b/ with another labial stop like /p/, e.g., “orb” vs “orp”). Replaced phonemes were drawn from the categories of stops (e.g., /b/ vs /p/), nasals (e.g., /m/ vs /n/), fricatives (e.g., /f/ vs /v/), affricates (e.g., /j/ in “edge” vs /e/ in “church”), liquids (e.g., /l/ vs /l/), and vowels. All nonwords were created so that replacement phonemes appeared approximately equally often in the beginning, middle, and end of the target and did not violate English phoneme combinations. The result was 400 word-nonword pairs matched in length and derivative word frequency (Thordike & Lorge, 1952).

Two stimulus lists (A and B) were created which both consisted of the same random order of word/nonword pairs, with each list containing one item from the pair. Stimulus items in List A were randomly assigned to appear as either a word or nonword and in either a high-noise or low-noise condition such that a quarter of the items ($n = 100$) appeared in each of the four conditions created by crossing noise and lexicality. List B was the mirror image of List A such that across the two lists each root target word appeared equally often in high and low noise and as a word and a nonword.

Stimuli (the 400 words and their corresponding nonwords) were recorded and edited using MacRecorder and SoundEdit Pro software for the Macintosh computer (MacroMind-Paracomp, 1990). They were recorded by a native female speaker of English with a sampling frequency of 22 Hz and edited such that each stimulus item had a duration of 1000 msec. These words and nonwords were spoken with natural inflection at approximately the same intensity. To verify this, the intensity of a sample of 10 stimuli drawn from across the experiment was measured using a Simpson Type S2A sound level meter (model 884) placed in the earpiece of Sony MDR-V1 headphones as the stimuli were played from the MacRecorder software on a Macintosh Quadra at the program’s default intensity, which was used to create the stimuli. Across the 10 items, intensities ranged from 85dB to 91dB ($M = 88.5, SD = 1.8$). One-second samples of white noise were created by SoundEdit Pro at program settings of 10% for low noise and 30% for high noise (the sound level meter showed that these were 70dB and 80dB, respectively). The target stimuli and segments of white noise were mixed using SoundEdit Pro by adding the intensities. Thus, the average signal-to-noise ratios for low and high noise, respectively, were 18.5dB and 8.5dB.

During the experiment, participants heard these stimuli through Sony MDR-V1 headphones directly from the audio port of a Macintosh IIci computer. Subjects individually adjusted the intensity to a comfortable listening level during practice, so that the listening level was free to vary while the signal-to-noise ratio was held constant (see Appendix A, Note 1).

Sentential contexts preceded the target item and were constructed to provide some semantic constraint for the word of the word-nonword pair. Sentences were between 10 and 15 syllables in length, and all were relatively simple in syntactic construction. Contexts were created so as to be of medium cloze value such that candidates were reduced but nevertheless numerous. They were visually presented to subjects on the computer screen in a large clear font (Geneva 18 bold).

Procedure

Participants were tested in a quiet room. Stimuli were presented on a computer using MacLaboratory software (Chute, 1994). The 400 stimulus words were recorded in two blocks of 200. Each block took 30 to 40 minutes to complete; each participant was required to take a break between blocks. Opportunities to halt the presentation of trials were programmed after every 20 trials to allow participants to take short breaks if needed. Participants resumed the presentation of trials by pressing the “word” key. During the auditory presentation of target words in the no-context condition, the screen was blank and remained so until the participant pressed a response key—either the “word” or “nonword” button on the keyboard. At this point, participants in the DF condition saw a “books” icon when they made correct detections and a “groceries” icon when they made correct rejections (these icons were counterbalanced across the two types of correct responses). When participants made incorrect responses (misses and false alarms), a third icon a “frown-face,” appeared. Participants in the SF condition saw the same icon (i.e., either the books or the groceries) for both types of correct responses and
the frown-face for incorrect responses. NF participants continued to see a blank screen (i.e., they had no visual presentations).

During an instructional period, participants were told that they would receive gift certificates that would vary in amount depending on how well they performed the task. Participants were also told that each correct response would be worth 4 cents toward the final gift certificate amount(s) which might be somewhere around $10 in total value. Those in the feedback conditions were told that the icons presented after each trial indicated whether they earned money toward a gift certificate (i.e., for the bookstore and/or grocery store) or not (i.e., the frown-face icon). Those in the DF condition were told the different meanings associated with the two “correct” icons; each presentation of the book icon meant they correctly identified a word as a word and that the grocery icon meant they correctly identified a nonword as a nonword (or vice versa for half of the participants in the DF condition). At the end of the second block of trials, the experimenter summed the total number of correct responses across both blocks and calculated the total value for the gift certificate(s) (rounded to the nearest dollar).

Both the blank screen and feedback presentations were of the same duration, 1000 msec, and the next trial began 2000 msec after the subject made a response. In the context condition, each context sentence was visually presented for 3000 msec and followed by the auditory stimulus. The feedback presentations (or lack thereof in the NF condition) were identical to those in the no-context condition. Prior to the experimental trials, participants began with a practice block of 20 trials (containing conditions representing all of the within-subject manipulations) in order to become comfortable with the procedure and to adjust the volume to a comfortable listening level.

RESULTS
Preliminary analysis showed that contrary to our expectations there were no systematic differences between the Same Feedback and Differential Feedback conditions. In order to simplify the presentation of our analyses, we have collapsed these two groups into a single Feedback (F) group to be contrasted with the No Feedback (NF) control.

Response Accuracy
A signal detection analysis was used to assess the effects of discriminability and bias in response accuracy. Estimates of discriminability (D) and bias (B) were obtained for each subject and the log transform of each was analyzed in a 2 (Age) × 2 (Context: C, NC) × 2 (Feedback: F, NF) × 2 (Noise Level: HN, LN) repeated-measures ANOVA in which Noise Level was a within-subject variable. Cell means for log D for the entire design are presented in Appendix B, and cell means for log B are presented in Appendix C (all analyses were conducted on the logs of D and B, but for ease of presentation, we refer to simply D and B in the following paragraphs).

Measures were screened for outliers. Within each age-feedback-block-noise cell (where block was either the first or second half of trials), values that were outside of 2.5 standard deviations around the mean were replaced with the cell mean. This resulted in a replacement of 2.9% of the data points; there were approximately an equal number of values replaced above and below the mean.

Discriminability (D).—With the exception of Feedback, $F < 1$, the main effects on D all reached significance, for the most part in a predictable fashion. D was higher for younger adults than for older adults ($M_Y = .67, M_O = .56$, $F(1,113) = 23.86, p < .001$; for the context condition than for the no-context condition ($M_C = .67, M_NC = .56$, $F(1,113) = 21.96, p < .001$; and for the low-noise condition than for the high-noise condition, ($M_{LN} = .75, M_{HN} = .48$, $F(1,113) = 398.10, p < .001$). The effects of noise were larger among the young than among the old, $F(1,113) = 15.25, p < .001$, and in the presence of context, $F(1,113) = 6.70, p < .01$, perhaps reflecting a floor effect in the high-noise condition.

Although it was surprising that feedback had no effect on overall D, there were some advantages among the older group conferred by having feedback that depended on the availability of context. In order to better understand the interplay among Age, Feedback, and Context, which was reflected in a marginally significant three-way interaction, $F(1,113) = 3.11, p < .08$, and in a marginal four-way interaction with Noise, $F(1,113) = 2.82, p < .1$, data for the context and no-context conditions were analyzed separately. In the no-context condition (but not in the context condition), the Age × Feedback × Noise interaction was significant, $F(1,56) = 5.47, p < .03$. This interaction reflects the fact that in the absence of context, feedback appeared to be somewhat deleterious to performance for the young in the low-noise condition and for the old in the high-noise condition (see Appendix B for cell means). In the context condition, however, feedback facilitated discrimination among the older group, but not among the young. This Age × Feedback interaction within the context condition, $F(1,57) = 4.22, p < .05$, is shown in the right-hand panel of Figure 1; in the left-hand panel is the nonsignificant interaction in the no-context condition, $F < 1$.

Thus, the positive effects of feedback were extremely restricted, namely only for the older group when sentential context was present. Otherwise, feedback seemed to have a negligible, or even slightly disruptive, effect on discrimination performance. We note that there is precedent in the literature for nonpositive effects of feedback. Schmidt and Bjork (1992) note that immediate feedback (that is presented on every trial as in the present case), although advantageous in terms of guiding the subject toward the right behavior, can also be detrimental to performance (particularly as measured in retention). One explanation for this phenomenon is that frequent feedback could block information processing activities that are important dur-

![Figure 1. Discriminability (log D) plotted as a function of age and feedback for the no-context (left panel) and context (right panel) conditions.](http://psychsocgerontology.oxfordjournals.org/Downloaded from http://psychsocgerontology.oxfordjournals.org)
ing the acquisition phase for acquiring the capability to produce effective performance at retention (p. 213). Even though feedback occasionally acted as a minor distracter, among the older listeners when it was presented in combination with sentential context it apparently provided the environmental support needed to augment performance.

Response Bias (B).—Generally speaking, participants in this experiment demonstrated a bias toward indicating stimuli were words, as shown by an overall positive value of log B that was statistically different from zero, which would indicate no bias (M = .12, with a 95% confidence interval of .085 to .164), F(1, 120) = 7.57, p < .001. Bias, however, was influenced by all three of our variables. Older adults were more likely than the young adults to indicate stimuli were words (Mf = .09, My = .18), F(1, 113) = 10.62, p < .01, suggesting that older listeners were more likely than the younger ones to construct meaning from the target. Thus, contrary to the suggestion that older adults are more conservative in signal detection tasks (Potash & Jones, 1977), older adults were more liberal in reporting a meaningful message from the ambiguous signal (Stine & Wingfield, 1987; Wingfield, Tun, & Rosen, 1995; see also Allen, 1990, for a demonstration of an age-related shift toward a more liberal criterion in memory for letter strings).

Bias was reduced by the presence of context (MC = .05, MNC = .22), F(1, 113) = 37.58, p < .001, so that having a sentential context for the ambiguous signal made it less likely that these would be perceived as meaningful items. Such a finding is consistent with word recognition models that posit a pattern-matching process between activated candidates, or prototypes, and the speech signal (e.g., Massaro, 1987). Apparently, the actual acoustic signal paled in comparison to the prototypes activated by the context prior to stimulus onset.

Although such a finding is at odds with some reports in the literature that context makes reconstruction of the speech signal more likely (Potter & Lombardi, 1990; Samuel, 1981), this difference is probably due to the highly available nature of the context in our experiment. Samuel (1981) showed that sentential context increased the bias toward reporting phonemic restoration for a degraded word in the middle of a spoken sentence, but such conditions would presumably place high demands on working memory, thereby enhancing the need for “regenerative” processes (Potter & Lombardi, 1990) of lexical repair. In our case, the written sentence was available for a full three seconds before the auditory signal was presented, presumably providing ample time for the activation of the phonological representations of possible candidates. In such a situation, context could enable critical analysis of the signal rather than promoting restoration of the expected phonemic pattern.

In contrast to the effects of sentential context, greater availability of intraword context, as given by noise level, increased bias. Bias was greater under low-noise conditions than it was under high-noise conditions, (MLN = .23, MHN = .04), F(1, 108) = 127.50, p < .001, so that regeneration was more likely when more phonological information of the word was available. Although sentential context appeared to reduce bias by enabling the comparison between acoustic input and the stored phonological representation, greater fidelity of the acoustic signal itself appeared to instantiate the phonetic pattern for the related word. This is consistent with a large literature showing that the probability of lexical access increases with the amount of phonemic information available (Lively et al., 1994). For example, Samuel (1981) showed that the probability of phonemic restoration was greater if the missing phoneme was later rather than earlier in the word, the interpretation being that an intact acoustic pattern early in the signal was sufficient to promote lexical access.

The Context X Noise interaction, F(1, 113) = 5.39, p < .05, was also significant. This was due to the greater effect of context in reducing bias when the level of noise was high, such that the combination of high noise and sentential context produced a slight bias for nonword responses (for LN, MNC = .29, MC = .17; for HN, MNC = .14, MC = -.06).

The paradoxical effects of noise and context on bias and their interaction are interesting in light of debates in the literature about the roles of phonological information and sentential constraints on word recognition. Our data are most consistent with interactive models in which both sentential context (Morton, 1969) and lexical context (i.e., phonological information; Marslen-Wilson & Tyler, 1980) activate candidates from the lexicon, and in which the pattern of the acoustic signal is matched against these activated prototypes (Massaro, 1987). The asymmetric effects of context and noise may suggest that the representations activated by lexical and sentential context, respectively, are by nature different, with lexical context driving restoration effects (Samuel, 1981) by feeding back to a prelexical level (Pitt, 1995), and sentential context providing a prototype for pattern-matching in the absence of prelexical feedback. Alternatively, the asymmetry may be accountable in terms of the relative on-line nature of lexical context and the relative offline nature of sentential context in the present experiment. In any case, the interaction between noise and context suggests that these processes may not be entirely independent.

Finally, there was a main effect of feedback, F(1, 113) = 18.50, p < .001, such that overall its presence markedly reduced bias (MNC = .05, MPA = .09). Nevertheless, this main effect appeared to be driven strictly by the NC condition, as reflected in a significant Context X Feedback interaction, F(1, 113) = 16.17, p < .001. As shown in Figure 2, context alone could virtually eliminate bias (within the context condition, MNC = .06, MP = .05); however without context, feedback had a substantial effect in reducing bias (within the no-context condition, MP = .32, MF = .11). This interaction did not vary as a function of age, F < 1, and Figure 2 shows this interaction for the younger participants (compare the left panels within the C and NC conditions) and for the older participants (compare the right panels within the C and NC conditions). Thus, for both young and old, sentential context was sufficient to bring judgments very close to the veridical; in the absence of context, bias was reduced by feedback, though not to the very low levels when context was available.

Figure 2. Bias (log B) plotted as a function of age and feedback for the no-context (left panel) and context (right panel) conditions.
Response Time

In addition to analyzing response accuracy, we also examined the effects of our variables on the time to execute correct judgments. First, there was no indication of a speed-accuracy trade-off (collapsed across all conditions, \( r(RT,D) = -0.08, p > .1 \), for all subjects; \( r(RT,D) = -0.02, p > .1 \), for young; \( r(RT,D) = -0.12, p > .1 \), for old; these correlations also failed to reach significance within each age-feedback-condition group).

Response times for words and nonwords were analyzed as a function of age, feedback, context, and noise. For words (i.e., hits), the only main effect to reach significance was noise, \( F(1,112) = 271.27, p < .001 \), indicating that hits were faster when the level of noise was low. The size of this effect was slightly greater when there was no feedback (for NF, \( M_{HN} = 1428, M_{MN} = 1546 \); for F, \( M_{HN} = 1450, M_{MN} = 1538 \)), as shown by a significant Noise \( \times \) Feedback interaction, \( F(1,112) = 5.36, p < .05 \). There was no main effect of age, \( F < 1 \), but there was a significant Age \( \times \) Noise interaction, \( F(1,112) = 3.98, p < .05 \), indicating that older adults were slightly faster than the young under low-noise conditions (for LN, \( M_L = 1455, M_O = 1430 \), but showed about the same response speed as the young under high-noise conditions (for HN, \( M_L = 1542, M_O = 1539 \)). For nonwords (i.e., correct rejections), there were facilitative effects of both feedback (\( M_{HN} = 1851, M_{MN} = 1652 \)), \( F(1,112) = 18.66, p < .001 \), and context (\( M_{HC} = 1758, M_{LC} = 1678 \)), \( F(1,112) = 4.16, p < .05 \). Additionally, correct rejections were faster under low levels of noise (\( M_{HN} = 1677, M_{MN} = 1759 \)), \( F(1,112) = 50.20, p < .001 \). The remaining effects and interactions were nonsignificant.

The fact that bias and response time for correct rejections both decreased with feedback made us wonder whether there might be a more direct connection. If the source of the bias were an effort toward reconstruction or regeneration, then subjects with greater bias scores would be expected to take longer to reject nonwords successfully when this process fails. To the extent that this effort toward reconstruction is advantageous in language processing, especially among elderly adults, as has been suggested in the literature (Stine & Wingfield, 1987; Wingfield et al., 1994; Wingfield & Stine-Morrow, in press), then those with high bias scores should be relatively faster in correctly detecting words—and this pattern should be exaggerated for older adults. The pattern of correlations provides some support for these ideas. Among older adults, bias was associated positively with time to reject nonwords correctly, \( r = .451, p < .001 \); bias also showed a weak negative association with time for correct detections, \( r = -.243, p = .06 \). Although the data from the younger adults showed a trend toward the same pattern, it was much weaker (for correct rejections, \( r = .271, p < .05 \); for correct detections, \( r = -.061, ns \). Table 1 shows these correlations when the data are broken down by either feedback condition (upper panel) or context condition (lower panel). These data show that the relationships between bias and response time are most pronounced in the absence of environmental support (i.e., either no feedback or no context; see Appendix A, Note 2). Thus, the data are suggestive of an age difference in the role of reconstruction in language processing.

Finally, we note that individual differences in processing capacity may also have contributed to response latencies. Consistent with other data showing a relationship between working memory span and speech understanding (Pichora-Fuller et al., 1995), older adults with higher forward and backward digit spans were faster at rejecting nonwords (the correlations for words (hits) for correct detections, \( r = .281, p < .01 \), and \( r = .271, p < .01 \), respectively, \( p < .05 \) for both). These relationships were weaker or nonsignificant for the correct detections of words, however \( r = -.239, p < .1 \), and \( r = -.186, p > .1 \). None of these relationships reached significance among the young, \( p > .1 \). Thus, these data suggest that the successful rejection of nonwords was especially capacity consuming for elderly listeners.

Table 1. Correlations Between Bias Scores and Response Time for Words and Nonwords (Correct Rejections)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Younger Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonwords Words</td>
<td>Nonwords Words</td>
</tr>
<tr>
<td>No Feedback</td>
<td>- .07</td>
<td>-.08</td>
</tr>
<tr>
<td>Feedback</td>
<td>.43†</td>
<td>- .07</td>
</tr>
<tr>
<td>No Context</td>
<td>.63***</td>
<td>- .32†</td>
</tr>
<tr>
<td>Context</td>
<td>-.07</td>
<td>- .32†</td>
</tr>
<tr>
<td>Overall</td>
<td>.27†</td>
<td>-.06</td>
</tr>
</tbody>
</table>

\( t_p < .10; * p < .05; **p < .01; ***p < .001 \).

DISCUSSION

The Effects of Age on Speech Discrimination

By examining lexical decisions as discriminated operands, we were able to separate out the effects of discriminability and bias in their contributions to age differences in speech processing. Consistent with other research examining age and speech processing, older adults in an auditory lexical decision task were found to show a reduced ability to make fine discriminations in speech input. Unlike research using other methodological paradigms, however, context was not reliable in mitigating these deficits (see Appendix A, Note 3). In hindsight, we realize that by crossing modalities with context sentences (which were presented visually) and target words (which were presented auditorially), we may have implemented a context manipulation that was suboptimal for elderly adults. Past research has shown that although younger adults can benefit from visual support of an auditory message, older adults may not (Stine, Wingfield, & Myers, 1990). Thus, the resource demands of having to integrate across the two modalities may have made it difficult for older adults to make full use of what we thought would be contextual support.

Contrary to some suggestions in the literature (Botwinick, 1984), we found no support at all for the contention that older adults are more cautious. To the contrary, in our experiment, they were more liberal in constructing meaning from an ambiguous signal. Although such reconstruction of the speech signal can, of course, generate error, it may nevertheless be reflective of a more general strategy of language processing that is advantageous in the long run. For example, older adults who make more reconstructive errors in the recall of syntactically scrambled strings nevertheless produce relatively higher recall scores (Stine & Wingfield, 1987). Similarly, in this case, older adults (but not younger adults) showed a tendency to recognize words more quickly when they were more biased toward word
judgments, particularly in the absence of external environmental support like feedback and context. To the extent that this is reflective of a more general mechanism, it could help account for the resilience of older adults’ performance in on-line speech processing in the face of cognitive declines (Wingfield, 1999; Wingfield & Stine-Morrow, in press).

**The Effects of Feedback on Speech Discrimination**

We expected that feedback would increase performance measurably for both young and old participants by enhancing the discriminability of the relationships between behavior and its consequences, thereby making the whole speech event more distinctive (Davison & Nevin, in press). Contrary to predictions of the discriminated operant model, we found no compelling evidence that increasing the distinctiveness of the feedback itself enhanced performance. It may be that the consequences themselves (icons for books vs groceries) were not different enough to produce an effect, but that remains to be tested in future research. Nevertheless, consistent with the discriminated operant model, we found limited effects of feedback on discrimination (among older adults when stimuli were embedded in sentential context), suggesting that consequences can enhance the distinctiveness of S-B-R relations.

It was surprising that feedback did not enhance discriminability at all among our younger subjects. Of course, it may be that our manipulation of outcome was not strong enough. Alternatively (or collaterally), it may have been that the superior sensory and processing systems of the younger adults enabled them to achieve perceptual constancy of the speech signal without the environmental support afforded by feedback. In fact, young adults showed their best levels of discriminability when context was available but feedback was not (see Figure 1).

By contrast, older adults, who often show evidence of diminished processing capacity in general (see Stine-Morrow & Miller, 1999, for a review) and word discrimination in particular (Sommers, 1996), took advantage of the feedback to enhance discriminability, showing their best performance when both semantic constraints and feedback were available (see Figure 1). Assuming that one function of context is to activate lexical candidates, in the present experiment the availability of context would have enabled a check of the veridicality of the match between the presented and canonical forms of a given phonological pattern. It may be that consequences were able to reduce the discrepancy between these acoustic representations among older listeners by shaping the canonical form to those produced by the speaker. Thus, the conditions under which older adults showed their highest levels of discriminability were when context activated an array of lexical candidates before acoustic information arrived and feedback shaped the accurate selection of candidates over a series of trials. We take this as support of the environmental support hypothesis (Craik & Jennings, 1992), and reject Baron and Surdy’s (1990) claim that aging brings an insensitivity to contingencies. In short, “an old dog can learn new tricks.”

Feedback served an additional function, however. Not only did feedback have limited effects on enhancing discriminability, but it also mitigated response bias toward meaningful reconstruction of the speech signal. This latter effect was true for young and old participants alike. Herein lies a paradox. Environmental support can both enhance discriminability and mitigate reconstruction, and yet the latter process, reconstruction, seems to serve an adaptive function among elderly adults, in fact, facilitating the speed of recognizing words. The answer seems to lie in the existence of boundary conditions on the usefulness of reconstructive processing. Clearly, elderly listeners demonstrated their highest levels of discriminability in the presence of environmental support. In the absence of this support, however, reconstructive processing is more likely among both young and old, but among the old, this seems to be related to enhanced word recognition.

**ACKNOWLEDGMENTS**

This research was supported by grants R15 AG11459 and R01 AG13935. We express our thanks to Phil Allen, Dan Morrow, and two anonymous reviewers for helpful comments on an earlier draft of this article, as well as to Barbara Herman and Jennifer Wright for assistance in testing research participants. Address correspondence to E. A. L. Stine-Morrow, Department of Psychology, Conant Hall, University of New Hampshire, Durham, NH 03824. E-mail: eals@christa.unh.edu

**REFERENCES**


AGE DIFFERENCES IN SPEECH DISCRIMINATION


Appendix A

Notes

1. Even though participants were offered the opportunity to adjust the intensity level of the stimulus, relatively few did so. Due to experimenter error, participants were started at “sound level 5” from the Macintosh operating system (approximately 94dB) in the no-context condition and “sound level 1” in the context condition (approximately 84dB). In the context condition, two elderly and no younger participants adjusted the intensity up to level 2; in the no-context condition, eight elderly and six younger participants adjusted the intensity downward to levels ranging from 4 to 2. There was no enough variability in the no-context condition to determine whether stimulus intensity affected D or B (recall that regardless of intensity, the signal-to-noise ratio was held constant). It did not, r(SL,D) = -.23, r(SL,B) = -.25, p > .1 for both.

2. One objection that could be raised is that because older adults had higher bias scores than the young, the age difference in correlations is simply a statistical artifact of differences in variability. However, bias scores can, of course, be negative so that zero is not necessarily a floor, and as Appendix C shows, there was no restriction of range among the young that would have prevented the manifestation of this relationship.

3. In fact, a more fine-grained analysis in which stage of practice was considered showed that age differences were mitigated only when the level of noise was high in the first half of the experiment, p < .01. It has been suggested that environmental support (Craik & Jennings, 1992), and context in particular (Wingfield et al., 1994), requires some measure of cognitive resources for their successful utilization. By the second block of 200 trials, participants may well have experienced some fatigue, perhaps depleting some measure of resources. Furthermore, older adults have been shown to take better advantage of context in relatively noisier environments (Cohen & Faulkner, 1983). The combination of these two factors may have accounted for our very limited demonstration of an age-related advantage in using contextual constraints.
### Appendix B. Cell Means and Standard Deviations for Discriminability (log D)

<table>
<thead>
<tr>
<th></th>
<th>No Context</th>
<th>Context</th>
<th></th>
<th>No Context</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Noise</td>
<td>High Noise</td>
<td>Low Noise</td>
<td>High Noise</td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>.70</td>
<td>(.12)</td>
<td>.48</td>
<td>(.10)</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>.62</td>
<td>(.19)</td>
<td>.47</td>
<td>(.15)</td>
<td>.68</td>
</tr>
</tbody>
</table>

### Appendix C. Cell Means and Standard Deviations for Bias (log B)

<table>
<thead>
<tr>
<th></th>
<th>No Context</th>
<th>Context</th>
<th></th>
<th>No Context</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Noise</td>
<td>High Noise</td>
<td>Low Noise</td>
<td>High Noise</td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>.35</td>
<td>(.18)</td>
<td>.16</td>
<td>(.29)</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>.13</td>
<td>(.13)</td>
<td>.01</td>
<td>(.09)</td>
<td>.12</td>
</tr>
</tbody>
</table>

Old

<table>
<thead>
<tr>
<th></th>
<th>No Context</th>
<th>Context</th>
<th></th>
<th>No Context</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Noise</td>
<td>High Noise</td>
<td>Low Noise</td>
<td>High Noise</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>.47</td>
<td>(.28)</td>
<td>.31</td>
<td>(.17)</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>.21</td>
<td>(.18)</td>
<td>.08</td>
<td>(.16)</td>
<td>.20</td>
</tr>
</tbody>
</table>