Lexical Neighborhood Density Effects on Spoken Word Recognition and Production in Healthy Aging

Vanessa Taler, Geoffrey P. Aaron, Lauren G. Steinmetz, and David B. Pisoni

We examined the effects of lexical competition and word frequency on spoken word recognition and production in healthy aging. Older (n = 16) and younger adults (n = 21) heard and repeated meaningful English sentences presented in the presence of multitalker babble at two signal-to-noise ratios, +10 and −3 dB. Each sentence contained three keywords of high or low word frequency and phonological neighborhood density (ND). Both participant groups responded less accurately to high- than low-ND stimuli; response latencies (from stimulus offset to response onset) were longer for high- than low-ND sentences, whereas response durations—time from response onset to response offset—were longer for low- than high-ND stimuli. ND effects were strongest for older adults in the most difficult conditions, and ND effects in accuracy were related to inhibitory function. The results suggest that the sentence repetition task described here taps the effects of lexical competition in both perception and production and that these effects are similar across the life span, but that accuracy in the lexical discrimination process is affected by declining inhibitory function in older adults.

Key Words: Aging—Lexical competition—Lexical discrimination—Neighborhood density—Speech production—Spoken word recognition

EVERYDAY conversation requires both comprehension of auditory input and production of spoken language. Both of these processes are influenced by linguistic characteristics of the words to be produced or understood; however, most research to date has focused on one or the other of these processes. In this article, we present a set of experimental materials and a task analysis that simultaneously tap perception and production processes, and discuss novel findings indicating that these materials reveal the differential effects of lexical competition on speech production and speech perception in both younger and older adults.

Lexical competition and spoken word recognition

Successful recognition of a spoken word requires the listener to match phonological input to the correct lexical entry in the mental lexicon. This entails identifying the perceived word from among tens of thousands of other word forms, some of which may be acoustically very similar to the target word. This process is known as lexical discrimination. An influential model of lexical discrimination is the Neighborhood Activation Model (NAM; Luce & Pisoni, 1998). According to the NAM, the process of spoken word recognition proceeds as follows: (a) perception of a spoken word activates the item’s lexical entry as well as those of words that are acoustically similar to the spoken word (lexical neighbors) and (b) the perceived word is selected from the activated lexical neighborhood, while competitors (i.e., neighbors) must be inhibited. The NAM defines lexical neighbors as words that can be formed from the target word through deletion, addition, or substitution of a single phoneme (Greenberg & Jenkins, 1964). For example, “can,” “scat,” and “at” are all lexical neighbors of the word “cat.” According to the NAM, two factors influence the discrimination and selection of a given word: its word frequency and the number of lexical neighbors it has, or its neighborhood density (ND). If a word has a high frequency of use in the language, it is easier to discriminate than a word of low frequency; likewise, a word with few lexical neighbors will be easier to discriminate than one with many lexical neighbors. Using three experimental paradigms—perceptual identification of words in noise, auditory lexical decision, and auditory word naming—Luce and Pisoni confirmed that young adults recognize isolated lexically easy (high frequency and/or low ND) words more quickly and accurately than isolated lexically difficult (low frequency and/or high ND) words.

An important aspect of the NAM is the assumption of both excitation (activation) and inhibition in lexical discrimination. Substantial evidence now exists for a decline in inhibitory functioning in healthy older adults (Hasher & Zacks, 1988). These inhibitory declines affect spoken word recognition in the older population: older adults show lower performance than younger adults on recognition of high-ND items, both when presented as isolated words, and in sentence contexts (low- and high-constraint; Sommers, 1996; Sommers & Danielson, 1999). Additionally, performance on an inhibitory task, the auditory Stroop task (Green & Barber, 1981, 1983; Jerger et al., 1993) has been found to correlate with spoken...
word recognition performance for high- but not low-ND items (Sommers & Danielson). Thus, declines occur in lexical discrimination in healthy aging, specifically for words that are more lexically difficult because they have many neighbors (i.e., there is increased competition among them). Declines in lexical discrimination result in greater difficulty understanding spoken language, which may explain in part the communication impairments that are consistently observed in older adults, above and beyond the language comprehension declines that would be predicted due to declining hearing function (Pichora-Fuller, 2003).

Lexical competition and language production

A small body of research also exists on the effects of phonological ND on language production. This research has been conducted largely independently of the perceptual research on phonological neighborhoods and spoken word recognition. Moreover, investigators have used different methodologies, principally an examination of the characteristics of tip of the tongue (TOT) states (Harley & Bown, 1998; Vitevitch & Sommers, 2003), speech errors (Vitevitch, 1997), picture naming (Newman & German, 2005; Vitevitch & Sommers), naming to open-ended sentences, and naming to category exemplars (Newman & German). This research has found that, in general, more phonological neighbors (i.e., higher ND) facilitate production in TOT and speech error tasks, resulting in fewer TOT states and speech errors (Harley & Bown; Vitevitch, 1997, 2002) in both younger and older adults (Vitevitch & Sommers). Conflicting results have been reported for naming: faster naming of high-ND words has been reported in both younger and older adults (e.g., Vitevitch, 2002; Vitevitch & Sommers), whereas high ND also has been reported to reduce naming accuracy, with similar effects observed in adults ranging in age from their 20s to their 70s (Newman & German).

Comparing speech perception and production

In sum, ND appears to facilitate production of words in both younger and older adults in TOT, speech error, and naming tasks (although cf. Newman & German, 2005). These findings are in contrast to ND effects in speech perception, in which high-ND words are more difficult to recognize. However, the differing methodologies used in these two lines of research make comparisons between these two processes difficult.

The present research aimed to fill this gap by using a sentence repetition task to assess both spoken word recognition and production. We used a modified version of the Veteran’s Affairs Sentence Test (VAST), originally developed by Bell and Wilson (2001). The VAST is an auditory sentence repetition task based on principles of the NAM. The original VAST sentences have been validated for intelligibility in a laboratory setting (Bell, 1996; Bell & Wilson; Lin, 2000); however, they are not controlled for target key word predictability, a factor that is known to be important in spoken word recognition studies (e.g., Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Sommers & Danielson, 1999). We thus modified the original stimuli to control for predictability.

The VAST was constructed to assess lexical discrimination using accuracy as a dependent variable (Bell & Wilson, 2001). However, we reasoned that both perception and production are required to successfully complete a sentence repetition task, and both processes can thus be examined simultaneously using this simple task. We measured both processes by examining temporally distinct phases of the participant’s responses: the perception phase, prior to the participant’s response (as measured by accuracy in repetition), and the production phase, in which the participant repeats the stimulus (as measured by response duration). We also measured response time—that is, the period between offset of the stimulus and onset of the response—that likely includes both perception and production processes, such as speech planning and organization. The approach used here has the advantage of holding participant characteristics, experimental stimuli, and testing environment constant. Thus, any differences observed in the effects of ND are not attributable to differences in task demands, stimuli, participants, or environment, but rather reflect the effects of ND on speech perception or production.

Methods

Participants

Participants were healthy young (n = 21) and older (n = 24) adults. Young adults were recruited from the Indiana University student population and through word of mouth. Older adults were recruited through the Neurology Clinic at the Indiana University School of Medicine, where they were being followed as healthy control participants, and through word of mouth. All participants were native English speakers with no neurological or psychiatric history. Demographic information and neuropsychological test scores are provided in Table 1 (8 older participants were excluded due to hearing loss, as discussed below; the information in Table 1 includes the remaining 37 participants). Participants were paid for their participation.

Audiometric Hearing Screening.—Screening for all participants was conducted in a quiet room. Pure tone air conduction thresholds were determined at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz using an audiometer (AMBCO Model 650A Audiometer, Tustin, CA), first in the right ear and then the left ear. The hearing screening was used as an inclusion criterion in order to minimize the possibility that group differences were due to hearing function. Participants were excluded if hearing loss exceeded 25 dB hearing level (HL) in the better ear at 500, 1000, or 2000 Hz or 35 dB HL at 4000 Hz (American National Standards Institute, 2004). Eight older participants had hearing loss that exceeded these criteria and were thus excluded from further analyses.
NEIGHBORHOOD EFFECTS IN AGING

Table 1. Demographic and Neuropsychological Information for Study Participants

<table>
<thead>
<tr>
<th></th>
<th>Young adults</th>
<th>Older adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mean ± SD)</td>
<td>(mean ± SD)</td>
</tr>
<tr>
<td>N</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>N by condition</td>
<td>11 at SNR of +10 dB</td>
<td>9 at SNR of +10 dB</td>
</tr>
<tr>
<td></td>
<td>10 at SNR of −3 dB</td>
<td>7 at SNR of −3 dB</td>
</tr>
<tr>
<td>Age (years)</td>
<td>22.19 ± 4.40</td>
<td>68.26 ± 7.39</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.10 ± 1.55</td>
<td>15.84 ± 3.00</td>
</tr>
<tr>
<td>Sex</td>
<td>5 men, 16 women</td>
<td>4 men, 12 women</td>
</tr>
<tr>
<td>Boston Naming Test (/60)**</td>
<td>54.33 ± 2.82</td>
<td>57.42 ± 3.24</td>
</tr>
<tr>
<td>Stroop—word naming*</td>
<td>107.7 ± 1.77</td>
<td>91.24 ± 27.61</td>
</tr>
<tr>
<td>Stroop—color naming***</td>
<td>80.57 ± 9.91</td>
<td>65.24 ± 14.55</td>
</tr>
<tr>
<td>Stroop—color-word naming***</td>
<td>54.10 ± 8.47</td>
<td>36.71 ± 9.29</td>
</tr>
<tr>
<td>Forward digit span (raw)</td>
<td>11.90 ± 2.70</td>
<td>11.58 ± 2.12</td>
</tr>
<tr>
<td>Backward digit span (raw)</td>
<td>9.00 ± 2.55</td>
<td>8.58 ± 2.09</td>
</tr>
<tr>
<td>MoCA (/30)*</td>
<td>28.90 ± 1.18</td>
<td>27.47 ± 2.04</td>
</tr>
</tbody>
</table>

Notes: BNT = Boston Naming Test; MoCA = Montreal Cognitive Assessment; SNR = signal-to-noise ratio.
*Group difference, p < .05.
**Group difference, p < .01.
***Group difference, p < .001.

Hearing thresholds for the remaining 37 participants are provided in Table 2.

Neuropsychological Testing.—All participants completed a short neuropsychological test battery comprising the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983), the Stroop Color and Word Test (Stroop, 1935), the forward and backwards digit span test from the Wechsler Memory Scale, 3rd edition (Wechsler, 1997), and the Montreal Cognitive Assessment (Nasreddine et al., 2005). Average scores by group are provided in Table 1.

Materials

Sentence Repetition Task.—Stimuli. The original VAST materials consist of 320 spoken sentences, each containing three target words distributed in four conditions: high frequency/high ND, high frequency/low ND, low frequency/high ND, and low frequency/low ND. All target words are monosyllabic and were rated as highly familiar in an independent norming study using students from Indiana University (Nusbaum, Pisoni, & Davis, 1984). Sentences were produced by a female native speaker of American English. Accurate speech recognition thresholds have been found with 8–10 sentences, and test–retest reliability varies by 2.5 dB, indicating that the test is reliable (Bell, 1996; Lin, 2000). In the present study, we used the original digital versions of the VAST sentences (i.e., they were not rerecorded).

Because, as mentioned earlier, the VAST stimuli were not controlled for word target predictability, we conducted a pretest that was designed as follows. Each stimulus sentence occurred three times in the pretest, once with each target word missing, for a total of 960 items in the pretest. The full-length pretest was divided into eight versions, and no sentence appeared more than once in any version. A total of 400 undergraduate college students at Indiana University completed the pretest for course credit (50 per list). We then calculated the average cloze probability (i.e., predictability) for each sentence by averaging the percentage of respondents who provided the correct word (i.e., the word that appears in the stimulus sentence) for each of the three target words. Lists of 40 sentences were then selected from the VAST compact disc database for each of the four conditions, 20 of which were of higher cloze probability and 20 of which were of lower cloze probability, for a total of 20 sentences (i.e., 60 target words) per condition (see Table 3 for an illustration of the conditions included in the experiment). We note that cloze probability was low across both conditions—in the higher cloze probability condition, average predictability was 0.14 (+/− 0.08), whereas in the lower cloze probability condition, average predictability was 0.01 (+/− 0.01). Target words in the two cloze probability lists were matched for word frequency, phonological ND, neighborhood frequency, number of higher frequency phonological neighbors, number of phonemes, and number of syllables, according to data from the English Lexicon Project (Balota et al., 2007).

Table 2. Hearing Thresholds (dB HL) for 500 to 4000 Hz by Age Group

<table>
<thead>
<tr>
<th></th>
<th>Younger adults</th>
<th>Oldeer adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mean ± SD)</td>
<td>(mean ± SD)</td>
</tr>
<tr>
<td>500 Hz</td>
<td>8.33 ± 6.18</td>
<td>9.38 ± 6.56</td>
</tr>
<tr>
<td>1000 Hz</td>
<td>7.50 ± 5.22</td>
<td>6.56 ± 7.47</td>
</tr>
<tr>
<td>2000 Hz*</td>
<td>3.06 ± 4.58</td>
<td>7.50 ± 7.30</td>
</tr>
<tr>
<td>4000 Hz**</td>
<td>7.22 ± 5.75</td>
<td>16.25 ± 10.08</td>
</tr>
</tbody>
</table>

Notes: Hearing thresholds are for the better ear at each frequency. Hearing thresholds were not available for three young adults, although all participants met the minimum threshold requirements stated in the text.
*Group difference, p < .05.
**Group difference, p < .01.

Table 3. Conditions Included in the Experiment and Sample Sentences

<table>
<thead>
<tr>
<th></th>
<th>Higher cloze probability</th>
<th>Lower cloze probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>High frequency, high ND</td>
<td>The first half of the test was hard.</td>
<td>The mold on the book made it stick.</td>
</tr>
<tr>
<td>High frequency, low ND</td>
<td>The point of the knife is too sharp.</td>
<td>The wives were tired and had many needs.</td>
</tr>
<tr>
<td>Low frequency, high ND</td>
<td>He will sob if you wreck his new bike.</td>
<td>Hurl a rotten peach at the witch.</td>
</tr>
<tr>
<td>Low frequency, low ND</td>
<td>He needed to greet the king in the castle.</td>
<td>Yawn as you lift about your badge.</td>
</tr>
</tbody>
</table>

Note: Target words are underlined. ND = neighborhood density.
Stimulus Presentation. Each sentence was presented in the presence of three-talker babble; all stimuli were presented to both ears. Participants were assigned to one of two conditions: babble set to a signal-to-noise ratio (SNR) of +10 dB or to an SNR of −3 dB. These SNRs were selected to elicit accuracy levels of approximately 95% and 70%, respectively, based on pilot data obtained in a pretest with six young normal-hearing adults. The VAST was administered using PsyScript 5.1d3 (Bates & D’Oliveiro, 2003) on a laptop computer (Apple PowerBook G4, Cupertino, CA). The two lists (higher and lower cloze probability) were presented as separate blocks in counterbalanced order, and participants were given a rest break between the two lists. Within each block, stimuli appeared in a different randomized order for each participant. Both signal and noise were presented to both ears over headphones (Beyerdynamic DT 770, Berlin, Germany) at a comfortable listening level. Participants wore a lapel microphone (Shure Microflex Condenser Mic, Evanston, IL) and were asked to listen to each sentence and simply repeat it back to the experimenter exactly as they heard it. The stimulus presentation was controlled by the experimenter, who pressed a button on the keyboard to present the next stimulus. The stimulus item and the participant’s response were recorded simultaneously on separate channels with a digital audiotape (Tascam DAT Recorder Model DA-P1, Tokyo, Japan) for later off-line scoring and acoustic analysis.

Data Analyses.—Response latencies for each stimulus were measured from the offset of the stimulus to the onset of the participant’s response (i.e., the moment that the participant began to repeat the sentence). Duration of each response was measured from the onset to the offset of the response using a digital waveform editor. To adjust for across-condition differences in stimulus duration, we calculated a normalized duration, which was defined as the length of time that the participant’s response lasted divided by the length of the stimulus.

In the accuracy analysis, data were entered into a mixed-model analysis of variance (ANOVA) with cloze probability (high/low), frequency (high/low), and ND (high/low) as within-participants factors and group (younger/older) and SNR (−3/+10) as between-participants factors. In the analyses of response time and duration, the higher and lower cloze probability conditions were collapsed due to low numbers of correct responses in the SNR −3 dB condition. Thus, data were entered into a mixed-model ANOVA with frequency (high/low) and ND (high/low) as within-participants factors and group (younger/older) and SNR (−3/+10) as between-participants factors. Significant interactions were decomposed with least square difference post hoc tests. For duration measures, two participants did not have enough accurate responses to be included in the analysis, and for response time measures, one participant was excluded for this reason.

For both response latency and duration measurements, two independent raters carried out separate measurements using a digital waveform editor, and the average of the two measurements was used in the analyses. Interrater reliability was 0.95 for latency measures and 0.94 for duration measures (calculated using an intraclass correlation analysis). Outliers were defined as any response latency or duration that was greater than 2.5 SDs from the mean for that participant and condition.

Results

Analysis 1: Accuracy

Average accuracy by group, condition, and SNR is presented in Figure 1. Accuracy was significantly higher overall for high- than low-frequency words, main effect of frequency, $F(1, 34) = 275.78$, $p < .001$, partial Eta squared = 0.89, and for low- than high-ND words, $F(1, 34) = 17.78$, $p < .001$, partial Eta squared = 0.34. Furthermore, accuracy was higher at an SNR of +10 dB than at an SNR of −3 dB, main effect of SNR, $F(1, 34) = 133.46$, $p < .001$, partial Eta squared = 0.80, and in higher than lower cloze probability stimuli, $F(1, 34) = 15.21$, $p < .001$, partial Eta squared = 0.31. Both frequency and ND effects were strongest at the lower SNR, interaction between frequency and SNR, $F(1, 34) = 103.32$, $p < .001$, partial Eta squared = 0.75 and interaction between ND and SNR, $F(1, 34) = 10.00$, $p < .01$, partial Eta squared = 0.23. ND effects were stronger in lower than higher cloze probability stimuli, interaction between ND and cloze probability, $F(1, 34) = 10.28$, $p < .01$, partial Eta squared = 0.23, and in low-frequency stimuli, interaction between ND and frequency, $F(1, 34) = 5.03$, $p < .05$, partial Eta squared = 0.13, particularly at the lower SNR, interaction between ND, frequency, and SNR, $F(1, 34) = 5.92$, $p < .05$, partial Eta squared = 0.15.

With respect to group differences, lower accuracy was observed for older than for younger adults overall, main effect of group, $F(1, 34) = 8.29$, $p < .01$, partial Eta squared = 0.20; group differences were greatest at an SNR of −3 dB, interaction between SNR and group, $F(1, 34) = 8.05$, $p < .01$, partial Eta squared = 0.19. ND effects were stronger in older than in younger adults, interaction between ND and group, $F(1, 34) = 5.24$, $p < .05$, partial Eta squared = 0.13, particularly at the lower SNR, interaction between ND, group, and SNR, $F(1, 34) = 5.40$, $p < .05$, partial Eta squared = 0.14. Finally, the strongest ND effects were observed in older adults for low-frequency stimuli at the lower SNR, interaction between frequency, ND, SNR, and group, $F(1, 34) = 7.71$, $p < .01$, partial Eta squared = 0.18.

Analysis 2: Response Latencies

Average response latencies by group and condition are presented in Figure 2. Response latencies were shorter to high- than low-frequency items, $F(1, 32) = 37.24$, $p < .001,$
Analysis 3: Response Duration

Average durations by group and condition are presented in Figure 3. Overall, response durations were shorter to high- than low-frequency items, $F(1, 32) = 8.57, p < .01$, partial Eta squared = 0.22, and to high- than low-ND items, $F(1, 31) = 8.43, p < .01$, partial Eta squared = 0.21. The strongest ND effects were in high-frequency items, interaction between ND and frequency, $F(1, 31) = 5.16, p < .05$, partial Eta squared = 0.14. Furthermore, durations were significantly longer for older than for younger adults (main effect of group, $F(1, 31) = 37.13, p < .001$, partial Eta squared = 0.55).

Analysis 4: Correlations With Neuropsychological Functions

To assess links between neuropsychological function and the effects of word frequency and ND, we computed a “frequency effect” score (average of high-frequency conditions subtracted from average of low-frequency conditions) and an “ND effect” score (average of most difficult ND conditions...
subtracted from average of easiest ND conditions, where “easiest” means high ND for response duration and low ND for accuracy and response time). These two scores were then correlated with the measures of neuropsychological function that previous research has suggested are potentially relevant to production and perception effects: Stroop color–word naming and forward and backward digit span. They were also correlated with hearing thresholds in the better ear at 500, 1000, 2000, and 4000 Hz (see Table 4). Because of SNR effects in response time and accuracy measures, the two SNR conditions were analyzed separately. At an SNR of −3 dB, a significant correlation was observed between the color–word naming condition in the Stroop test and the ND effect in accuracy ($r = -0.60$, $p < .05$) and between hearing threshold in the better ear at 2000 Hz and the ND effect in accuracy ($r = 0.54$, $p < .05$), indicating that lower inhibitory function is associated with larger differences between low and high ND on accuracy measures when listening conditions are more

HH = high frequency, high neighborhood density
HL = high frequency, low neighborhood density
LH = low frequency, high neighborhood density
LL = low frequency, low neighborhood density

Figure 2. Average normalized duration by signal-to-noise ratio (SNR), group, and condition. Error bars represent standard error. Normalized duration was measured as the time from response onset to response offset divided by the length of the stimulus sentence.

Figure 3. Average response time by signal-to-noise ratio (SNR), group, and condition. Error bars represent standard error. Response time was measured as the time from stimulus offset to response onset.
difficult (at the lower −3 dB SNR). Furthermore, greater short-term memory capacity (as measured by forward digit span) was associated with a larger word frequency effect in response latency at an SNR of −3 dB (r = 0.53, p < .05). Finally, at an SNR of +10 dB, a borderline significant correlation was observed between the word frequency effect in accuracy and forward digit span (r = 0.44, p = .051), suggesting that greater short-term memory capacity may also be associated with a larger word frequency effect in spoken word identification.

**DISCUSSION**

In the present experiment, we found significant age differences overall, with lower accuracy and longer response durations for older than younger adults, a result that is consistent with previous research findings (Smith, Wasowicz, & Preston, 1987; Sommers, 1996). ND exerted a significant effect on all measures. Higher accuracy and shorter response times were observed for low- than high-ND words, whereas shorter response durations were observed for high- than low-ND words. These findings are consistent with previous research indicating that having many lexical neighbors causes greater competition in spoken word recognition (Luce & Pisoni, 1998), but facilitates language production (Harley & Bown, 1998; Vitevitch, 1997, 2002).

The pattern of differential effects of ND on speech production and perception measures was observed in both participant groups, indicating that lexical competition continues to exert similar effects throughout the life span. However, the effect of ND on accuracy was stronger in older than in younger adults, particularly when listening conditions were most challenging: for low-frequency stimuli, at the lower SNR, and in sentences with lower cloze probability. For accuracy measures, a significant negative correlation was also observed between Stroop color–word naming performance and the size of the ND effect at the lower SNR: that is, lower

inhibitory function is associated with lower performance on high-ND words under more difficult listening conditions, consistent with previous research (Sommers & Danielson, 1999). The ND effect in accuracy also correlated significantly with hearing threshold in the better ear at 2000 Hz, suggesting that declines in auditory acuity also play a role in older adults’ lower performance in recognizing more difficult words under difficult listening conditions.

Word frequency effects, in contrast, appeared to be related to short-term memory capacity, as indicated by the associations observed between performance on forward digit span and the size of the frequency effect. This finding may be related to the fact that the methodology used in the present study required participants to repeat words from short meaningful subspan sentences, in contrast to previous research focusing on generation of spoken language (e.g., Harley & Bown, 1998; Vitevitch & Sommers, 2003). The present results suggest that, at least in the sentence repetition task used here, declines in short-term memory capacity may result in more difficulty in retaining low-frequency words in working memory while processing subsequent material and/or in planning speech (as measured by response latency). In the present study, the sentences were designed to be similar in length and memory demands; furthermore, the older adults who took part in our study were cognitively intact and performed similarly to younger adults on tests of short-term and working memory. These factors limit our ability to assess the effects of short-term memory decline on lexical competition and frequency effects; these effects should be further explored in future research using stimuli that specifically manipulate short-term/working memory demands in repetition, as well as with participants with a documented impairment in short-term memory.

For both accuracy and response latency measures, effect sizes for word frequency were larger than those for ND, whereas for duration measures, they were equivalent. Effect
sizes were largest in accuracy (word frequency: 0.89, ND: 0.34), followed by reaction time (word frequency: 0.54, ND: 0.24), and were smallest in duration measures (word frequency: 0.22, ND: 0.21), suggesting that stimulus characteristics affect lexical discrimination more in perception than in production, both for younger and older adults, at least using the measures included in the present study. The differences in word frequency and ND effect sizes obtained with the different measures suggest that word frequency experts a greater effect than ND on speech perception but not speech production.

The present study focused primarily on adults who fall into the young–old range, and the older adults in the present study reflect a select high-functioning group with relatively good hearing who were able to complete the task accurately. As such, the results reported here may not be generalizable to hearing-impaired and lower functioning older adults or to middle–old and old–old adults—although the results are consistent with previous research with older samples (e.g., Sommers, 1996; Sommers & Danielson, 1999). We also note that the present results should be interpreted as preliminary because of the small sample size. In future research, we plan to extend our findings to a larger and more heterogeneous sample, including older adults at risk for dementia. Research with this population would shed further light on the impact of neuropsychological function on spoken word recognition and the ways in which these processes may change with cognitive impairment. Previous research suggests that patients with mild Alzheimer’s disease (AD) have greater difficulty than healthy older adults in discriminating lexically difficult words, possibly due to declines in processing speed and/or inhibitory function in lexical discrimination (Sommers, 1998). In the Sommers (1998) study, participants were required to repeat a single word embedded in a carrier sentence (“Please say the word ____ for me”). However, the effects of lexical competition in sentence context have yet to be explored in this clinical population; given differences in the use of sentence context in comprehension in healthy aging and AD (e.g., Schwartz, Federmeier, Van Petten, Salmon, & Kutas, 2003), lexical competition effects may be expected to interact with sentence context in these populations.

A second avenue for further research is the use of a broader, more comprehensive neuropsychological battery that includes more sensitive measures of neuropsychological function, as well as measures of speech discrimination, in addition to the hearing screen used in the present study. It should also be noted that previous studies of the effects of lexical competition on spoken word recognition in aging have used white noise (Sommers, 1996) or speech-shaped noise (Sommers & Danielson, 1999). In contrast, we presented stimuli in the presence of multitalker babble, which creates greater competition and produces more degradation in speech intelligibility. Researchers have suggested that multitalker babble has greater ecological validity than other types of noise, such as speech-shaped noise or white noise (e.g., Kalikow, Stevens, & Elliot, 1977). However, multitalker babble also contains temporal dips and can thus produce local reductions in masking; moreover, it should be noted that Kalikow and colleagues described 12-talker babble, whereas the present study used 3-talker babble, in which dips are likely even greater. Age differences have been found in the ability to take advantage of these local dips (Peters, Moore, & Baer, 1998); it is thus possible that the type of background noise may interact with the lexical competition and age effects. Although the results reported in the present study are consistent with previous findings, future research should further explore this possibility in greater depth using the stimulus materials employed here.

The present study presents a novel approach to examining the effects of lexical competition in both production and perception using the same experimental task. Use of the same task controls for differences in task demands and experimental conditions that have made it difficult to link the findings relating to lexical competition in spoken word recognition and language production. The approach used here allowed us to differentiate between different stages of speech production and perception using accuracy as well as response duration measures, revealing contrasting effects of ND in the two measures. Accuracy measures proved the most sensitive to age differences, suggesting that age differences in lexical competition effects occur primarily at the level of successful identification of the stimulus rather than processing time required to identify and/or reproduce the item.

Previous research on effects of lexical competition on language production in younger and older adults has utilized primarily single-word responses, such as elicitation of TOT states or naming tasks. In the present study, in contrast, participants were required to repeat short meaningful sentences. Despite these methodological differences, a similar effect of ND was observed for both younger and older adults, with high ND facilitating speech production, as measured by response duration. These findings suggest the possibility that ND effects on production are not related to retrieval or speech planning demands per se—because these effects are not observed in accuracy or response latency measures—but rather may have their locus at the point of articulation and phonetic implementation. Recent work by Kemper, Schmalzried, Herman, Leedahl, and Mohankumar (2009) suggests that older adults may alter their speaking rate to compensate for increased task demands; that is, the present results may reflect changes in strategic processing due to the greater processing demands engendered by high-competition items in these sentences. This issue should be further explored using tasks specifically designed to isolate the stages of language production.

The findings reported here have several implications for understanding declines in everyday communication in healthy older adults, which contribute to declines in quality...
of life (Lubinski, 1991; Nussbaum, 2007). Similar effects of lexical competition were observed for the older and younger adults in the present study, although older adults were less accurate overall than younger adults in repeating high-competition words, particularly under difficult listening conditions. The significant correlations between performance on a spoken word recognition task and measures of inhibition and short-term memory abilities point to possible underlying cognitive changes that may drive declines in language comprehension. We are currently exploring the use of this sentence repetition task with patients diagnosed with mild cognitive impairment or probable Alzheimer’s disease to determine whether they differ from healthy older adults.

In summary, the sentence repetition task employed in this study can be used to assess the effects of lexical competition in both perception and production. Although the effects of word frequency and ND were found to be similar across the life span, accuracy in lexical discrimination was affected by declines in inhibitory function and short-term memory capacity in older adults.

**FUNDING**

This research was supported by funds from NIDCD T32 Training Grant DC-00012 and NIDCD Research Grant DC-00111. V.T. was supported by a postdoctoral fellowship from Fonds de Recherche en Santé du Québec.

**ACKNOWLEDGMENTS**

We thank Andrew Kirk and Adrienne Roman for their help with the acoustic measurements and Luis Hernandez for valuable technical advice. We are grateful to three anonymous reviewers for valuable comments on an earlier version of the manuscript.

**CORRESPONDENCE**

Address correspondence to Vanessa Taler, MA, PhD, School of Psychology, University of Ottawa, 200 Lees Ave., Ottawa, Ontario, Canada K1N 6N5. Email: vtaler@uottawa.ca.

**REFERENCES**


