Older listeners' decreased flexibility in adjusting to changes in speech signal reliability

Laurence Bruggeman[1] and Esther Janse[2]

[1]The MARCS Institute, University of Western Sydney, Australia; [2]Centre for Language Studies and Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands
l.bruggeman@uws.edu.au; e.janse@let.ru.nl

ABSTRACT

Under noise or speech reductions, young adult listeners flexibly adjust the parameters of lexical activation and competition to allow for speech signal unreliability. Consequently, mismatches in the input are treated more leniently such that lexical candidates are not immediately deactivated. Using eyetracking, we assessed whether this modulation of recognition dynamics also occurs for older listeners. Dutch participants (aged 60+) heard Dutch sentences containing a critical word while viewing displays of four line drawings. The name of one picture shared either onset or rhyme with the critical word (i.e., was a phonological competitor). Sentences were either clear and noise-free, or had several phonemes replaced by bursts of noise. A larger preference for onset competitors than for rhyme competitors was observed in both clear and noise conditions; performance did not alter across condition. This suggests that dynamic adjustment of spoken-word recognition parameters in response to noise is less available to older listeners.

Keywords: Spoken-word recognition, aging, processing dynamics, hearing loss.

1. INTRODUCTION

When listeners try to understand speech, they have to segment a continuous stream of speech input into separate words. This is not a trivial task, as it is not always clear where one word ends and the next one begins and because longer words can contain embeddings of shorter words. As the speech signal unfolds, words that overlap with parts of the speech signal are activated in the listener's mind until they are no longer supported by the speech signal and can be ruled out as viable candidates. The words that 'win' this lexical competition are those words that account best for the speech input without leaving any phonemes unaccounted for [10].

In normal listening conditions, words that overlap with the start of the spoken word (onset competitors, e.g., *circus*-*circle*) compete more strongly for recognition than words that overlap with the end of the spoken word (rhyme competitors, e.g., *cent*-*tent*). Studies using the visual world paradigm, in which participants' eye movements are recorded as they hear speech while viewing visual displays, have found that onset competitors typically attract more looks than unrelated distractor pictures [1]. Rhyme competitors also attract more looks than unrelated distractors but these effects are much smaller and occur later in time [1, 12]. Under adverse listening conditions, however (e.g., casually articulated speech or noise), younger adults adjust these competition processes [4, 11]. Consequently, mismatch between a lexical candidate and the incoming speech signal no longer necessarily leads to immediate de-activation of the candidate word; onset competitors compete *less* strongly, and rhyme competitors compete *more* strongly than in ideal listening conditions. This is explained as listener adjustments for the decreased reliability of the speech signal.

Participants in [4] and [11], as indeed in most studies of speech perception, were undergraduates. This flexibility in adjustment of competitor evaluation in word recognition is potentially one of the pillars of the robustness of speech perception under difficult listening conditions. Extensive research on speech perception in aging listeners, however, has shown that in adverse listening conditions, older listeners cope less well than younger listeners and that the disparity cannot be explained fully by hearing loss [5, 6]. In particular, inhibitory capacities have been argued to decline with age, which may influence the extent to which older adults are able to suppress lexical competitors [15, 18]. Like younger adults, older listeners show a fairly strong onset-competitor effect and a rhyme-competitor effect that is smaller than the onset-competitor effect [e.g., 2, 14] but their ability to adjust competitor relationships under adverse conditions has not been investigated.

In this study, we assessed whether native Dutch listeners, aged 60 years and over, adjust lexical competition when the speech signal is occasionally interrupted by bursts of noise. The saccadic motor system appears to be largely unaffected by aging [13], so we conducted an eyetracking experiment based closely on [11]. We did not use their exact methodology, but instead adapted it in light of the characteristics of our participant population. Older participants typically exhibit more individual differences than younger participants. We therefore changed the presence of noise bursts in the speech signal from a between-subjects variable to a within-subjects variable, i.e., all participants were presented both with sentences from the baseline (noise-free) condition and with sentences from the noise condition. This approach reduces the variance due to between-subject variability.[1]

If older listeners flexibly adjust competitor evaluation in response to the decreased reliability of the speech input, we expect a weaker onset-competitor effect and a stronger rhyme-competitor effect when noise is added to the speech signal.

2. METHOD

2.1. Participants

Twenty-two participants (11 males) from the subject pool of the MPI for Psycholinguistics in Nijmegen, the Netherlands, were paid for their participation in this study. Seven additional participants' results were excluded due to calibration difficulties. All participants were native speakers of Dutch, aged 62–85 years (*M*=69.8, *SD*=6.5), with normal or corrected-to-normal vision. None of the participants wore hearing aids in their daily life. Pure-tone air conduction thresholds were determined for all participants. The mean threshold for the better ear (averaged over 0.5, 1, and 2 kHz) was 22.6 dB HL (range: 3.3-43.3, *SD* = 12.8). High-frequency thresholds for the better ear (averaged over 4, 6 and 8 kHz) ranged from 3.3-70.0 dB HL (*M* = 38.3, *SD* = 24.2). Informed consent was obtained prior to the start of the experiment.

2.2. Stimulus materials

Stimuli were based on [11] and consisted of 120 recorded Dutch sentences each containing a critical word. Sentences were constructed in such a way that the critical word was not easily predictable (e.g., *Het zag eruit als een* <u>*paspoort*</u>*, maar de tekst op de voorkant klopte niet.*, "It looked like a <u>passport</u> but the text on the front was not correct.") and were spoken by a female native speaker of Dutch. They were read out with neutral intonation and the speaker was unaware of the presence or identity of any of the critical words. Each sentence was paired with a visual display containing four black-and-white line drawings. Using a so-called 'target absent' design [7], the critical words were not represented by any of the drawings in the two experimental conditions. Instead, the visual displays contained one phonological competitor for the critical word and three distractors that were phonologically and semantically unrelated. In the onset-competitor condition, the phonological competitor drawing depicted a word that overlapped at onset with the critical word (e.g., for the critical word *paspoort*, 'passport', the onset competitor was *paspop*, 'tailor's dummy'). In the rhyme-competitor condition, the competitor drawing depicted a word that had a rhyme overlap with the critical word and only differed in its first phoneme (e.g., for the critical word *honing*, 'honey', the rhyme competitor was *koning*, 'king'). In addition to these two experimental conditions of 40 sentences each, there were 40 filler sentences, for which the visual displays contained a picture of the critical word and three unrelated distractor pictures. In all conditions, competitor and distractor pictures were counter-balanced across four fixed positions on the screen. The sentences were recorded in a sound-attenuated booth at a sampling rate of 44.1 kHz using Adobe Audition. Subsequent selection, measurement and editing of the auditory stimuli was carried out using Praat [3]. Two versions were created of each sentence. One version consisted of the original recording and this sentence was used in the baseline condition. For the noise condition, a second version was created in which between two and four separate phonemes throughout each sentence were replaced with bursts of noise. As no effect of noise position was found by [11], noise bursts replaced word-initial phonemes in half of the sentences and word-medial phonemes in the other half of the sentences. Importantly, the bursts of noise were never inserted in the critical word, nor in the two words preceding and following the critical word. The same radio noises were used as in [11]. The duration of each noise burst was adjusted individually, so that each burst replaced exactly one phoneme. The mean noise duration was 81.2 ms (range: 11-214 ms, SD = 33.8 ms), well above the gap detection thresholds reported for older listeners [8, 16, 17]. The loudness of each noise burst was adjusted so that it corresponded to 80% of the average intensity (in dB) of the sentence it was inserted in. In each condition (onset-competitor, rhyme-competitor, and filler) there were 14 sentences containing two noise substitutions and 13 each with three and four substitutions. In each sentence, bursts were evenly divided over the sentence fragment that preceded the critical word and the fragment following it.

2.3. Procedure

Participants were tested individually in a sound-attenuated booth. They were seated in front of a computer screen at a viewing distance of 95 cm, with their head held in a fixed position by means of a chin and forehead rest. Participants' eye movements were recorded at a sampling rate of 1000

Hz (monocular) using an Eyelink 1000 Tower Mount system (SR Research, Ltd.). Auditory stimuli were presented over Sennheiser HD201 headphones at a loud but comfortable level, kept constant for all participants.

Before the start of the experiment, the eyetracker was calibrated and validated using a 9-point calibration grid. After every five trials, an automatic drift correction was carried out and, if required, calibration was repeated. Following [11], participants were not given an explicit task, other than to listen to the sentences and to not take their eyes off the screen. At the start of each trial, a fixation cross was displayed in the centre of the screen. Participants were instructed to look at this cross until it disappeared. After this, visual displays were shown for 1s before the start of each sentence.

All participants were presented with all 120 sentence-display pairs in two blocks. The first block always contained sentences from the baseline (noise-free) condition, the second block consisted only of sentences from the noise condition. Items were counterbalanced across blocks and each participant was presented with a different randomisation of the stimulus list. There was no break between the baseline and the noise block and participants were not informed about the presence of noise in the second phase of the experiment.

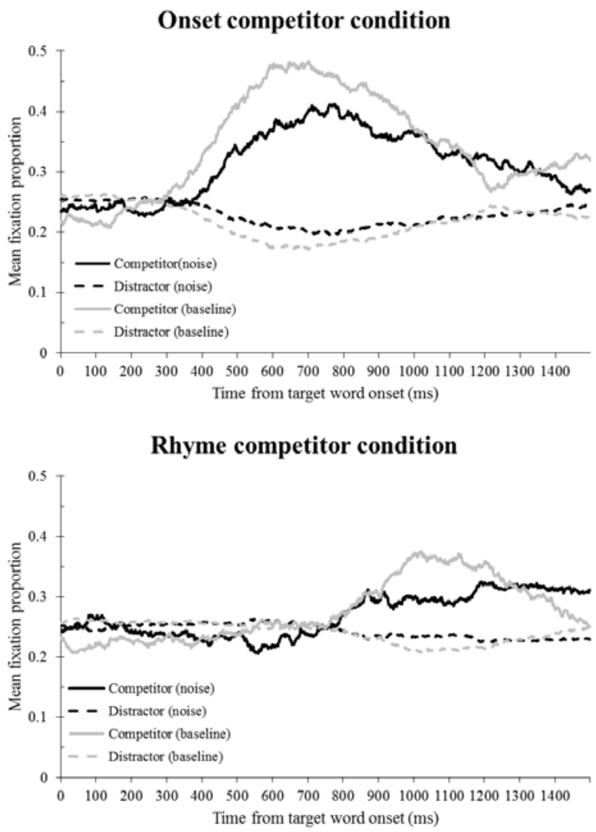
Upon completion of the eyetracking task, participants filled in a background questionnaire.

3. RESULTS

Figure 1 shows the average fixation proportions to the competitors and distractors from the onset of the target word for the onset- (top) and the rhyme-competitor condition (bottom). Distractor fixations were divided by three to account for the fact that each visual display contained three distractors and only one competitor picture.

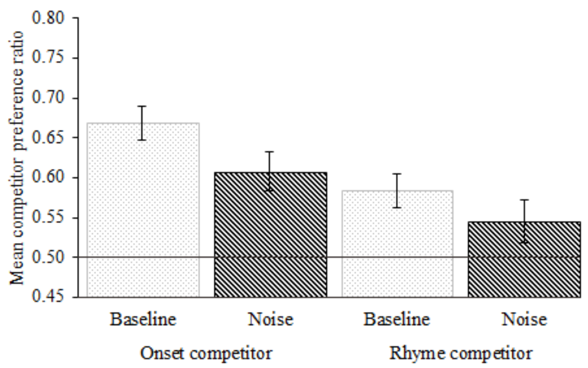
As it is generally assumed that it takes around 200 ms to initiate an eye movement [9], the earliest time windows used for analysis in visual world studies, including [11], begin at 200 ms after target word onset. Because of our older participant population we have chosen to analyse windows that start later in time. Competitor preference ratios were therefore computed over a 600 ms time interval, starting at 300 ms after critical word onset for the onset-competitor condition and at 800 ms for the rhyme-competitor condition (see Fig. 2). For each type of competitor, this was done by dividing the total number of fixations to the competitor by the sum of competitor fixations and distractor fixations. The number of distractor fixations was divided by three to account for the fact that for every competitor picture there were three distractors in every display.

Figure 1: Mean fixation proportions from target word onset



In a one-sample two-tailed *t* test by participants (1) and by items (2), competitor preference ratios were compared to 0.5. In the baseline condition, both onset [$M_1 = 0.67$, $t_1(21) = 8.04$, $p < .001$; $M_2 = 0.64$, $t_2(39) = 6.21$, $p < .001$] and rhyme [$M_1 = 0.58$, $t_1(21) = 3.98$, $p = .001$; $M_2 = 0.58$, $t_2(39) = 3.25$, $p = .002$] competitors were fixated significantly more than distractors. In the noise condition, onset competitors were fixated significantly more than distractors [$M_1 = 0.61$, $t_1(21) = 4.48$, $p < .001$; $M_2 = 0.57$, $t_2(39) = 2.57$, $p = .014$], whereas rhyme competitors were not [$M_1 = 0.55$, $t_1(21) = 1.73$, $p = .099$; $M_2 = 0.54$, $t_2(39) = 1.44$, $p = .157$].

Figure 2: Mean competitor preference ratios per condition for the 600 ms analysis window



A 2x2 repeated-measures analysis of variance (ANOVA) of fixation ratios in the critical time window was conducted by participants (F_1) and items (F_2), with noise type (baseline and noise) and competitor type (onset and rhyme) as the within-subject factors. This showed a main effect of noise condition [$F_1(1,18) = 4.44$, $p = .047$; $F_2(1,78) = 5.44$, $p = .022$]. We also found a main effect of competitor type (onset competitors were fixated more than rhyme competitors) but while this effect is significant across participants, it misses significance over items [$F_1(1,21) = 13.21$, $p = .002$; $F_2(1,78) = 2.99$, $p = .088$]. No trace of a noise condition by competitor type interaction was found [$F_1(1,18) = 0.30$, $p = .591$; $F_2(1,78) = 0.24$, $p = .625$].

Further, as participants varied in hearing acuity from normal-hearing to mild-to-moderate hearing loss, we checked for links between participants' hearing thresholds and their preference for competitor pictures over distractors. No correlation was found between participants' competitor preference ratios (in any of the four within-subject conditions) and their pure-tone air conduction thresholds averaged over 0.5, 1 and 2 kHz and averaged over 2, 4 and 8 kHz (with *r*-values ranging from -.17 to .13).

4. DISCUSSION

We investigated whether older listeners adjust the parameters of lexical activation and competition when the reliability of the speech signal is decreased by occasional bursts of noise. If these listeners' performance mirrored that of the listeners in [11], we predicted a weaker onset-competitor and a stronger rhyme-competitor effect in the noise condition than in the baseline (noise-free) condition,

First, our results in the baseline condition confirmed previous findings of strong onset-competitor effects and smaller rhyme-competitor effects in young and older adults [e.g., 1, 2]. This suggests normal efficiency of speech processing in older listeners. However, in the baseline condition, mean fixation proportions to both onset and rhyme competitors reached higher peaks (0.48 and 0.37 respectively) than those found by [11] for younger adults in the baseline condition. Fixation proportions in [11] peak around 0.34 for onset competitors and around 0.24 for rhyme competitors. This may indicate that, even in noise-free listening conditions, older listeners are more cautious than younger adults in eliminating competitors as potential lexical candidates. Alternatively, it could be the result of the decrease in inhibitory capacities that is generally associated with aging: older listeners may experience more difficulties suppressing competitors than younger adults.

In the noise condition, onset competitors attracted fewer looks than in the baseline condition. This replicates the findings by [11] and could indicate that listeners adjust the parameters of lexical activation and competition. However, we also found a smaller competitor preference for rhyme competitors in noise, which speaks against such an adjustment. Listeners' increased uncertainty about the speech signal was expected to lead to a larger preference for rhyme competitors. Even though older adults were clearly affected by the noise bursts, they did not adjust their processing dynamics. This might be linked to the fact that rhyme competitors already attracted a high proportion of looks from older listeners in the baseline condition. In order to compensate for age-related deficits, older adults may employ dynamic listening strategies in noise-free situations that are similar to those used by younger adults to adjust to noisy listening conditions. When the speech signal deteriorates, older listeners could already be operating at capacity and may therefore not adjust further to the changing listening conditions.

As we did not exclude participants with poorer than normal hearing thresholds, about half of all participants suffered from mild to moderate hearing loss. Even though the difference between the results of this study and that by [11] may (partly) relate to hearing differences between age groups, there was no correlation between individual hearing thresholds and older participants' competitor preferences.

In sum, contrary to previous findings for younger adult listeners, we do not find conclusive evidence that suggests older listeners adjust the parameters of lexical activation and competition when the speech signal becomes less reliable due to the presence of noise. If it is the case that older listeners already make this type of adjustments while processing speech in normal listening conditions, this appears to leave them without resources to fall back on when listening conditions become difficult.

5. REFERENCES

[1] Allopenna, P.D., Magnuson, J.S., Tanenhaus, M.K. 1998. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38(4), 419-439.

[2] Ben-David, B.M., Chambers, C.G., Daneman, M., Pichora-Fuller, M.K., Reingold, E.M., Schneider, B.A. 2011. Effects of aging and noise on real-time spoken word recognition: evidence from eye movements. *Journal of Speech, Language & Hearing Research*, 54(1), 243-262.

[3] Boersma, P., Weenink, D., Praat: doing phonetics by computer, 2013.

[4] Brouwer, S., Mitterer, H., Huettig, F. 2012. Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539-571.

[5] Frisina, D.R., Frisina, R.D. 1997. Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hearing Research*, 106(1–2), 95-104.

[6] Helfer, K.S., Staub, A. 2014. Competing speech perception in older and younger adults: Behavioral and eye-movement evidence. *Ear and Hearing*, 35(2), 161-170.

[7] Huettig, F., Altmann, G.T.M. 2005. Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), B23-B32.

[8] Lister, J.J., Roberts, R.A., Lister, F.L. 2011. An adaptive clinical test of temporal resolution: Age effects. *International Journal of Audiology*, 50(6), 367-374.

[9] Matin, E., Shao, K.C., Boff, K.R. 1993. Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53(4), 372-380.

[10] McQueen, J.M. 2007, Eight questions about spoken-word recognition. In: Gaskell, M.G. (ed), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, 37-53.

[11] McQueen, J.M., Huettig, F. 2012. Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America*, 131(1), 509-517.

[12] McQueen, J.M., Viebahn, M.C. 2007. Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, 60(5), 661-671.

[13] Pratt, J., Dodd, M., Welsh, T. 2006. Growing Older Does Not Always Mean Moving Slower: Examining Aging and the Saccadic Motor System. *Journal of Motor Behavior*, 38(5), 373-382.

[14] Revill, K.P., Spieler, D.H. 2012. The effect of lexical frequency on spoken word recognition in young and older listeners. *Psychology and Aging*, 27(1), 80-87.

[15] Robert, C., Mathey, S. 2007. Aging and Lexical Inhibition: The Effect of Orthographic Neighborhood Frequency in Young and Older Adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62(6), P340-P342.

[16] Snell, K.B. 1997. Age-related changes in temporal gap detection. *The Journal of the Acoustical Society of America*, 101(4), 2214-2220.

[17] Snell, K.B., Frisina, D.R. 2000. Relationships among age-related differences in gap detection and word recognition. *The Journal of the Acoustical Society of America*, 107(3), 1615-1626.

[18] Sommers, M.S., Danielson, S.M. 1999. Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, 14(3), 458-472.

---

[1] In a pilot study reported at the 41st Australasian Experimental Psychology Conference, Brisbane, April 2014, we replicated the exact methodology used in [11] but with older listeners. Results were entirely parallel to those discussed in this paper, so here we report only the methodologically improved, subsequent experiment.