LISTENING TO FAST SPEECH: AGING AND SENTENCE CONTEXT

Esther Janse[1,2], Majoke van der Werff[1] & Hugo Quené[1]

Utrecht institute of Linguistics OTS[1] and Max Planck institute for Psycholinguistics Nijmegen[2]
esther.janse@let.uu.nl

ABSTRACT

In this study we investigated to what extent a meaningful sentence context facilitates spoken word processing in young and older listeners if listening is made taxing by time-compressing the speech. Even though elderly listeners have been shown to benefit more from sentence context in difficult listening conditions than young listeners, time compression of speech may interfere with semantic comprehension, particularly in older listeners because of cognitive slowing. The results of a target detection experiment showed that, unlike young listeners who showed facilitation by context at both rates, elderly listeners showed context facilitation at the intermediate, but not at the fastest rate. This suggests that semantic interpretation lags behind target identification.

Keywords: spoken word processing, aging, fast speech, hearing, time compression

1. INTRODUCTION

As people age, they may suffer from age-related impairments in hearing. These impairments include a progressive reduction in absolute hearing sensitivity, mainly for higher frequencies, and a decline in discrimination abilities (such as temporal discrimination) [1]. Additionally, aging involves cognitive decline: older adults may become slower in terms of information processing [2]. The combination of such auditory and cognitive problems can make speech perception demanding for older adults, particularly in background noise or when speech is fast.

To overcome their initial speech decoding problems, elderly listeners may rely more strongly on semantic context. Age differences are smaller when listeners have to report last words of high-predictability sentences than of low-predictability sentences [3,4,5]. If bottom-up acoustic information does not suffice to uniquely identify words, top-down support from context may increase activation of contextually consistent targets, so as to increase the relative difference between the target and its competitors.

However, in contrast to these context results obtained with offline methods, recent neuroimaging studies (using the N400) have shown that the effect of a constraining sentence context is smaller in older than in young listeners [6]. Similarly, older people were shown to be less affected when targets are unexpected in a given context than young people. This suggests that older adults do not use context predictively [7]. Thus, age differences in contextual benefit may be dependent on the listener task or processing stage that the experiment taps into. If context is not used predictively, it may still play a role at later stages.

In this study, we investigated speech processing in young and elderly listeners and their use of sentence context. Speech processing was made more taxing by time-compressing the speech to 1.5 times and to 2 times its normal rate. Older adults are affected much more by an increase in speed than young adults [8,9], which can be attributed to both hearing deficits and cognitive slowing.

Because of these contradictory results on contextual benefit, we were interested to find out how easily pre-assigned target words could be detected in normal sentences and in semantically unpredictable sentences at two rates of speech. If meaningful context helps to identify spoken words in difficult listening conditions, target detection time should be affected *more* by the presence of meaningful context at the faster rate than at the intermediate rate (Hypothesis 1). Alternatively, if cognitive slowing is an important factor in elderly listeners' problems with time-compressed speech, elderly listeners may not have had the opportunity to build a semantic representation of the sentence. Given that older subjects, contrary to young subjects, did not use context predictively (even) at normal speech rate, building a semantic representation seems a relatively slow process for them. One would then predict a *smaller* effect of meaningful context at the faster rate than at the moderately fast rate, particularly for the elderly listeners (Hypothesis 2). Target detection time (and accuracy) in the different sentence conditions was compared between elderly and young listeners to test these conflicting hypotheses. By also assessing each subject's hearing and processing speed, we wanted to investigate the relation between these two factors and their ability to process fast speech.

2. METHOD

A target detection study was set up in which the target was presented visually before a spoken sentence was played to the subjects. Stimuli, design, procedure and participants are given below.

2.1. Stimuli

As targets 96 disyllabic nouns were selected (48 with initial stress and 48 with final stress). They were embedded in three contexts: normal meaningful sentences, semantically unpredictable sentences (SUS), and in word lists (the reason for including this latter condition is irrelevant to the present study). Targets were not highly predictable in the normal sentences (literal word-by-word translation, e.g., 'At the table the youths were with a guitar busy'). The position of the target item varied between sentences. For each meaningful sentence, a semantically unpredictable counterpart was constructed which was syntactically well-formed but which made no overall sense (e.g., 'At the chicken the spoon has round a guitar dreamt'). The position of the target, syntactic construction and overall sentence length were equal in each normal - SUS sentence pair. In the word-list context, 5, 6, or 7 words that had each been uttered in isolation were spliced together, separated by 400 ms silent intervals, such that the target was always the 5th item in test trials. In addition to the 96 test trials (rotated over conditions), there were a number of practice trials, plus 48 overlap and 48 miss trials in which the sentence contained a word that overlapped phonemically with the target, or in which the sentence did not contain the target.

A male native speaker of Dutch with a clear speaking style was asked to read all the sentence materials and isolated words at a normal rate. Recordings were made with a Sennheiser microphone in a sound-treated booth: the speech files were then downsampled to 32 kHz. In order to achieve a similar speech rate, phrasing and intonation contour across the two sentence contexts, the speaker read each SUS sentence directly after he had read its normal counterpart. Targets always carried sentence accent. Mean duration of the target word was 440 ms in the normal context (SD=56), 452 ms (SD=52) in the SUS context, and 584 ms in the isolated context (SD=84). An ANOVA showed that context had a significant effect on target duration ($F(2,285)=145.5$, $p<.001$). Post-hoc tests showed that target duration did not differ between the two sentence contexts ($p>.1$), but that the isolated condition differed significantly from both sentence contexts ($p<.001$ in both comparisons).

The speech materials were presented in two time-compressed conditions: compressed to 67% and to 50% of the original duration (these rates of 1.5 and 2 times normal rate were determined on the basis of a pilot study). The audio files were uniformly time-compressed using PSOLA as implemented in PRAAT (www.praat.org). The resulting speech is unnatural because speakers speed up in a non-uniform way [10].

2.2. Design and procedure

The three contexts (normal sentence, SUS, and word-list contexts) and the two compression rates made six conditions. To avoid target repetition within an experimental session, the 96 targets were rotated over the six conditions on six different stimulus lists (according to a Latin square design).

On each list, the six conditions were presented in separate trial blocks in a fixed order: first the normal sentence condition, then the SUS condition, and then the word lists were presented at 1.5 times normal rate. Then the three context conditions (in the same order) were presented at 2 x normal rate. Each block was preceded by at least seven practice trials containing hits, misses and overlaps (to ensure that initial adaptation to the fastest rate had taken place before the test trials began, the first 2x normal rate block had more practice trials). Within each block, the order of presentation of test and filler trials was randomised for each participant.

Participants were tested either in their homes or in the institute's speech lab. They were seated at a table in front of the computer screen, wearing closed headphones (Beyer Dynamic DT770). The targets were displayed visually in a large font 800 ms prior to sentence onset and remained visible during the sentence. Participants were asked to press the button of a button box as quickly as possible once they detected the assigned target (without sacrificing accuracy). The audio was played to them at a comfortable listening level. Following sentence offset, there was a 2 second interval during which listeners could still give their response. After 2.5 s, the next trial started.

2.3. Participants

18 young and 18 elderly subjects were recruited for this study (3 of each age group assigned to each list). They were all native speakers of Dutch. The young listeners (mainly university students) had a mean age of 22 years (SD=2.9). The elderly listeners, who were mainly recruited in a home for the elderly, had a mean age of 82 (range 70-97, SD=8.6). They had no major health problems that might interfere with their performance in this study (such as stroke, Parkinson's disease, or signs of dementia as judged by a speech therapist (2nd author).

Pure-tone air conduction thresholds were measured in dB (at 0.25, 0.5, 1, 2, 4 and 8 kHz) using a screening audiometer (Maico ST20). Overall hearing acuity was expressed as the average of hearing thresholds at 1, 2 and 4 kHz for the better ear: mean acuity was 2.7 dB for the young group (SD=3.2) and 32.9 dB for the elderly group (SD=15.4).

Additionally, all subjects carried out the Digit Symbol substitution subscale of the WAIS intelligence test [11]. This test is a measure of information processing speed and measures the number of symbols that can be recoded in a period of 90 seconds and, as a baseline, the number of symbols a participant can simply copy in 90 s. Mean coding speed was 40 symbols in 90 s for the elderly (SD=15) and 70 for the young subjects (SD=7). Mean copying speed was 80 for the elderly (SD=32) and 145 for the young subjects (SD=17). Age differences remained after correction for the motor component (by subtracting the time needed to copy this number of symbols from total processing speed): $t(34)=5.2$, $p<.001$.

3. RESULTS

Because this paper focuses on context facilitation in connected speech and thus on target detection in meaningful versus semantically unpredictable sentences, the target detection results of the word-list condition are not discussed here. First, as an indication of task difficulty, false detection rates (collapsed over overlap and miss conditions) are given in Table 1 for both speech rates and for the two context conditions. The relatively low false detection rates for both age groups showed that listeners did not press the button whenever something sounded like the target.

Second, results for accurate detection of targets in the different conditions is given in Table 2. Detection was only counted as accurate if the subject had pressed the button after target onset.

Table 1: False detection rates (%) in both age groups.

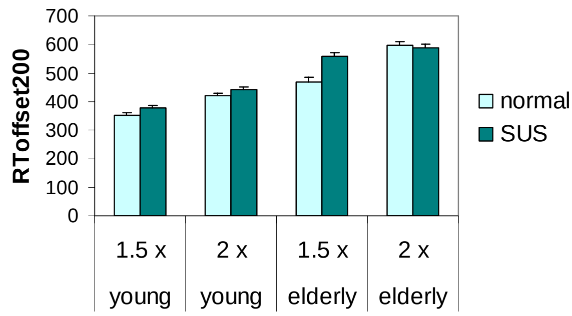
		Young	Elderly
1.5 x normal rate	Normal context	5	8
	SUS context	2	5
2 x normal rate	Normal context	4	5
	SUS context	3	3

Table 2: Correct detection rates in two test conditions in both age groups.

		Young	Elderly
1.5 x normal rate	Normal context	100	100
	SUS context	100	99
2 x normal rate	Normal context	100	94
	SUS context	99	95

Response times were measured from target offset. This was done to account for the fact that target duration was different in the two rate conditions and to account for the fact that targets were slightly longer in the SUS context. To avoid negative response times, however, a point 200 ms before target offset was chosen. In Fig. 1, mean detection times (from 200 ms before target offset) are plotted for the different contexts, rates, and age groups.

Figure 1: Mean target detection times (in ms, with standard errors) in two contexts for the two age groups.



Target detection times were logtransformed to normalise the data and entered into ANOVAs (by subjects and items) to test the effects of Age Group, Rate, and Context.

Response times of the elderly subjects were significantly longer than those of the young subjects (Age Group effect: $F_1(1,34)=26.6$, $p<.001$); $F_2(1,94)=383.3$, $p<.001$). RTs to speech at the fastest rate were longer (relative to target offset) than at the moderately fast rate (Rate effect: $F_1(1,34)=51.7$, $p<.001$; $F_2(1,94)=239.3$, $p<.001$). There was no interaction between Age and Rate ($F_1(1,34)<1$, *n.s.*; $F_2(1,94)<1$, *n.s.*), indicating that both age groups took longer to respond at the faster rate. Fig. 1 shows that target detection was generally facilitated by meaningful context in the normal sentence condition, which is indicated by a significant effect of Context (F_1(1,34)=18.6, *p*<.001; F_2(1,94)=27.7, *p*<.001). The interaction between Rate and Context was also significant (F_1(1,34)=9.6, *p*=.004; F_2(1,94)=16.1, *p*<.001), which indicates that the context facilitation effect differed for the two rates. Clearly, this was most notable for the elderly listeners, which is supported by a three-way interaction between Age, Rate and Context (F_1(1,34)=6.5, *p*=.015; F_2(1,94)=21.8, *p*<.001). The Age*Context interaction was not significant (F_1(1,34)<1, *n.s.*; F_2(1,94)=1.2, *n.s.*).

Pairwise comparisons (with Bonferroni adjustment for multiple comparisons) showed that the young listeners showed context facilitation at both rates of speech (*p*<.05 by subjects and items): the context effect even seemed slightly more consistent at the fastest rate, which is in line with the compensatory use of context as in Hypothesis 1. The elderly listeners, however, showed clear context facilitation at the moderately fast rate, but not at the fastest rate. Elderly listeners' detection accuracy was still high at the fastest rate, but the speech rate was too high for sentence comprehension to keep up. This result is in line with Hypothesis 2: semantic comprehension lags behind target identification.

To investigate the contribution of hearing acuity and processing speed to the elderly subjects' difficulty with the fast rates, we investigated the RT and correct detection rate data in linear mixed effects models with the experimental factors and hearing acuity and processing speed as predictors. There was a significant effect of elderly listeners' hearing acuity on their correct detection rate (*p*<0.01). The corrected Digit-Symbol Substitution speed measure did not significantly affect correct detection rate (nor RT). Thus, even though hearing and cognitive decline are assumed to play a role in elderly listeners' difficulty with fast speech, the cognitive measure did not predict performance here.

4. CONCLUSION

Both young and elderly listeners can still reliably detect pre-assigned words in speech that is played back at twice the normal rate. However, building a higher-level meaning representation may be a relatively slow process for the elderly. Time compression was shown to interfere with the immediate use of context during spoken-word processing. Context cannot help to resolve the competition between word candidates if comprehension lags behind.

The data analyses showed that age-related hearing impairment contributes to elderly listeners' difficulty with fast speech processing. Poorer perception of the sentence context will clearly reduce its potential facilitatory effect. However, even though the cognitive speed measure did not predict performance, cognitive slowing may interfere with the use of context as a means to compensate for poorer hearing, such that impaired hearing and cognitive slowing together yield a cumulative snowball effect. Further research is required to disentangle the relative contributions of hearing and cognitive factors to speech processing problems in the elderly.

5. REFERENCES

[1] Willott, J.F. 1991. *Aging and the auditory system: Anatomy, Physiology and Psychophysics.* San Diego: Thompson.

[2] Salthouse, T.A. 1996. The processing speed theory of adult age differences in cognition. *Psychological Review* 103, 403-428.

[3] Wingfield, A., Aberdeen, J.S., Stine, E.A. 1991. Word onset gating and linguistic context in spoken word recognition by young and elderly adults. *Journal of Gerontology* 46, 127-129.

[4] Pichora-Fuller, M.K., Schneider, B., Daneman, M. 1995. How young and old listeners listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593-608.

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

[6] Federmeier, K.D., Kutas, M. 2005. Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology* 42, 133-141.

[7] Federmeier, K.D., McLennan, D.B., De Ochoa, E., Kutas, M. 2002. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. *Psychophysiology* 39, 133-146.

[8] Gordon-Salant, S., Fitzgibbons, P.J. 1993. Temporal factors and speech recognition performance in young and elderly listeners. *J. Speech Hearing Res.* 36, 1276-1285.

[9] Wingfield, A., Tun, P.A., Koh, C.K, Rosen, M.J. 1999. Regaining lost time: adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging* 14, 380-389.

[10] Janse, E. 2004. Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication* 42, 155-173.

[11] Wechsler, D. 2000. *Wechsler Adult Intelligence Scale WAIS-III-NL.* Nederlandstalige bewerking: J. M. Uterwijk (ed). Lisse: Swets & Zeitlinger.