Memory For Speech Under Difficult Listening Conditions

Master's Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Interdepartmental Program in Neuroscience
Professor Arthur Wingfield, Advisor

In Partial Fulfillment
of the Requirements for the Degree

Master of Science
in
Neuroscience

by
Apoorva Arora

May 2017
ACKNOWLEDGEMENTS

The author would like to express her gratitude to Prof. Arthur Wingfield for his guidance, sagacious insight and encouragement throughout the project. The author would also like to thank Lana Penn, Nicole Amichetti, Alex Kinney and other members of the Memory and Cognition Lab for their help during the project.
ABSTRACT

Memory For Speech Under Difficult Listening Conditions

A thesis presented to the Interdepartmental Program in Neuroscience

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Apoorva Arora

Memory for speech relies on the processing of the speech signal as it is received. With aging, working memory and processing speed decline making comprehension of speech, and hence its memory, difficult. Older adults overcome this challenge by utilizing their preserved linguistic knowledge as long as the total by processing burden for speech is below a threshold of resource overload. An experiment is presented here in which the processing burden for comprehension of spoken sentences in increased by increasing the speech rate and adding syntactic complexity. The recall accuracy for these sentences is compared between young and older adults to examine the possible compensatory mechanisms employed by older adults in the face of such challenges in their everyday lives.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................................................................................ iii

ABSTRACT ................................................................................................................................................................................................ iv

LIST OF TABLES .......................................................................................................................................................................................................................................................... vi

LIST OF FIGURES ................................................................................................................................................................................................................................................ vii

INTRODUCTION ........................................................................................................................................................................................................................................... 1

Effects of Syntactic Complexity ......................................................................................................................................................................................... 2

Time Compression ......................................................................................................................................................................................................................... 4

METHODOLOGY ........................................................................................................................................................................................................................................ 6

Participants ............................................................................................................................................................................................................................... 6

Stimuli .............................................................................................................................................................................................................................. 7

Audibility Check .................................................................................................................................................................................................................. 8

Procedure ......................................................................................................................................................................................................................... 9

Scoring of the responses ........................................................................................................................................................................................................ 10

RESULTS .............................................................................................................................................................................................................................. 11

DISCUSSION ........................................................................................................................................................................................................................... 15

REFERENCES ...................................................................................................................................................................................................................... 18
Table 1. Examples of sentences being used as stimuli in this study.
LIST OF FIGURES

Figure 1. Recall Accuracy of young adults and older adults for the stimulus sentences.
INTRODUCTION

Comprehension of speech, a task that most people think is automatic, is actually quite complex. Successful comprehension of spoken language requires two levels of processing – one based on the sensory processing of the acoustic signal and the other based on cognitive processes such as comprehension of meaning using semantic information stored in memory. For the latter process, speech itself has to be stored in memory as it is being received. As the spoken discourse progresses, the segment of speech stored in memory is replaced the newer parts of spoken discourse (Jarvella, 1979). This piece of information in memory is then matched with the linguistic context for successful recognition.

If either of the two processes – sensory processing or cognitive processing – fails, such as in the case of hearing loss that interferes at the sensory level, or in the case of aging, where declines at the cognitive level are common, language comprehension breaks down. Adult aging leads to declines in episodic memory (Wingfield & Kahana, 2002), working memory (Salthouse, 1994) and processing speed (Salthouse, 1996). In addition, there is commonly an age-related decline in hearing acuity, particularly in the higher frequencies that support speech perception (Morell, Gordon-Salant, Pearson, Brant & Fozard, 1996). As a result, comprehension of speech becomes challenging for many older adults as these two levels of processing interact in an essential way.

In case of older adults with mild hearing loss, we do not see a catastrophic failure of comprehension or recall of speech. Rather, we see a reduced accuracy, saved in large measure by
the preserved linguistic knowledge. This can be seen in their reliance on the context for recall of what has been heard which is evident in the greater accuracy of recall for meaningful sentences rather than unrelated word-strings (Wingfield & Stine-Morrow, 2000).

The positive effects of preserved linguistic knowledge in older adults help to balance out the negative effects of decline in hearing acuity and cognitive resources that results in a more successful comprehension of speech than would be predicted. However, this compensation fails when the burden for processing the speech signal increases such as when speech is masked by background noise or it is presented at a rapid rate (Gordon-Salant & Fitzgibbons, 1993) or with a complex syntactic structure (DeCaro, Peelle, Grossman & Wingfield, 2016).

All of these situations tax both the perceptual level and cognitive level processes. If the processing demands increase, cognitive resources are diverted away from storage of speech stimuli in memory (Wingfield, Tun, & McCoy, 2005). One can then predict that under these circumstances, the memory for speech, measured by the accuracy of recall, would be affected and this effect would be more pronounced for older adults than for the young adults.

In the experiment presented here, I examine the accuracy of recall of older adults in comparison to young adults for sentences when the processing challenge was increased by presenting sentences with a complex syntax and high speech rates.

**Effects of Syntactic Complexity**

Processing challenge arises if the sentences become longer and/or are syntactically complex, i.e. they do not have the canonical *Subject-Verb-Object* form (Just & Carpenter, 1992). This *Subject-Verb-Object* construction is commonly used and referred to as the Subject Relative structure. For example, in the sentence ‘Girls that teach boys are nice’, the subject ‘girls’ is
followed by the verb ‘teach’ and then the object ‘boys’. A more challenging structure is the Object Relative structure in which the sentence departs from the canonical *Subject-Verb-Object* form. For example, in the sentence ‘Boys that girls teach are nice’, the object ‘boys’ is followed by the subject ‘girls’ and then the verb ‘teach’. Studies in the past have reliably shown that more errors are produced both in recall and comprehension of sentences with an object relative structure. Furthermore, the performance of older adults in recall and comprehension of these sentences is worse than that of young adults.

The processing challenge can be increased further by separating the subject and its object by a prepositional phrase. This separation can be long or short, with the longer the separation, the more difficult the challenge. To exemplify, in the sentence ‘Girls on long city walks that teach boys are nice’, the subject ‘girls’ is separated from its action ‘teach’ by the prepositional phrase ‘on long city walks’. The separation here between the subject+verb is a long one. This is opposed to the short separation in this form of the same sentence: Girls that teach boys on long city walks are nice. The sentences with insertion of these prepositional phrases, especially when they increase the Subject-Verb distance are more challenging as they require greater working memory resources. If working memory deficits exist or cognitive resources are insufficient, the processing challenge increases and such sentences become more difficult to recall or comprehend (DeCarro *et al.*, 2016).

In this experiment, sentences have been presented in four syntactic structures: Subject-Relative with a long separation, Subject-Relative with a short separation, Object-Relative with a long separation & Object-Relative with a short separation. In doing so, the processing challenge should have increased allowing us to observe age-related decline in cognitive resources on verbatim recall of these sentences.
**Time Compression**

In addition to the introduction of syntactic complexity with the inclusion of the prepositional phrase, the processing challenge has been further increased by presenting the sentences at variable rapid speech rates. Studies have shown that with impairments of processing speed and working memory, as seen in older adults, the recall of time-compressed or rapid speech is adversely affected. This is thought to be because less time is available for processing speech from memory if the speech signal is being received very rapidly (Wingfield, Poon, Lombardi, & Lowe, 1985).

This rapid speech has been produced by an acoustic manipulation—Time Compression of Speech. Time-compression of the stimuli sentences is done by the sampling method that preserves the relative temporal patterning of the speech which is essential for speech processing. In this method, small sections of the speech signal are removed to generate a compressed signal that has a reduced playing time compared to the original speech signal. For instance, to compress a signal to 75% or $\frac{3}{4}$th of its original playing time, every fourth piece of data is removed from the signal by the computer software. For speech recorded at 22kHz, there would be 22,000 pieces of data in 1 second of the signal and so, removal of these very small pieces of data would not lead to a noticeable loss of speech information. In addition, the pitch contour of the original signal would be preserved. Moreover, as this piece of data is removed equally from both the speech and silent periods, the relative temporal pattern of natural speech is maintained. The resulting signal produces speech in less time than normal but is still intelligible (Wingfield, Tun & Rosen, 1995).

In the following experiment, I have used three different speech rates, normal or compression ratio of 100 which is the unmodified normal speech rate, compression ratio of 65
which is 65% of the original playing time and compression ratio of 55 which is 55% of the
original playing time.
METHODOLOGY

Participants

Participants were 16 older adults (3 males and 13 females) in the age group of 65-85 years \((M = 75\) years) who were monolingual native English speakers. All participants reported themselves to be in good health with no history of the use of hearing aids or of speech disorders, Parkinson’s disease, stroke or any other neuropathological condition that would interfere with their ability to carry out the tasks. All the participants in this group were well educated with an average of 18.25 years (S.D. = 3.04) of formal education. They also showed a good knowledge of vocabulary, with a mean score of 15.75 (S.D. = 2.32) on the Shipley Vocabulary Test (Zachary, 1991). The hearing acuity of all the participants was measured using a GSI 61 Clinical Audiometer (Grason Stadler Inc., Madison, WI) in a sound-attenuated testing room. The better ear Pure-Tone Average (PTA), averaged over 0.5, 1, 2 and 4 kHz, ranged from 7.5 to 41.25 \((M = 25.39\, \text{dB HL}; \, S.D. = 9.66)\). As is typical for this age group and degree of hearing loss none of the participants reported themselves to be a regular user of hearing aids (Fischer, Cruickshanks, Wiley, Klein, Klein, & Tweed, 2011) and all testing was conducted unaided.

For purposes of comparison a group of 16 young adults (ranging from 18 to 24 years of age was also included. They had age-normal hearing acuity with mean better ear PTA ranged from -2.5 to 7.5 \((M = 5.54\, \text{dB HL}; \, S.D. = 4.99)\). The young adults had completed fewer years of formal education \((\text{mean} = 13.68\, \text{years}, \, S.D. = 1.1)\) at the time of testing compared to the older
adult participants, \( t(30) = 2.954, p = 0.006 \). As is common (Verhaeghen, 2003) they also had lower scores on the Shipley Vocabulary Test with the mean score for the group being 13.75 (S.D. = 1.39). Similar to the older adults, all the young adults also reported themselves to be healthy, monolingual native English speakers.

**Stimuli**

**Stimulus Sentences**: The stimuli were a subset of 576 sentences taken from DeCaro et al. (2016). All the sentences were constructed such that a male (e.g. monk, king, father etc.) or a female (e.g. nun, queen, mother etc.) agent performed an action and their female or male counterparts respectively were the receivers of the action. 288 sentences expressed the meaning with in the Subject-Relative (SR) structure and 288 with the more complex Object Relative (OR) structure. These sentences with these two syntactic structures were further divided into 144 sentences in which a 4-word adjectival phrase was inserted into a sentence such that there was a short separation between the agent of the action and the action being described (short separation) and 144 sentences in which the position of the adjectival phrase produced a longer separation between the agent and the action (long separation). Examples of these four sentence types are shown in Table 1.

All sentences were recorded by a female American English speaker in natural intonation at a normal speech rate (150 wpm) on Sound Studio software (V 2.2.4), Macromedia, Inc., San Francisco, CA. After recording, the sentence sound files were normalized within each sentence and across all sentences to the root mean square intensity using the PRAAT software (v6.0.21, Institute of Phonetic Sciences, University of Amsterdam).
**Table 1:** Examples of sentences being used as stimuli in this study. Words in red are the agents and words in blue are the actions.

<table>
<thead>
<tr>
<th>Syntactic Structure</th>
<th>Separation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Relative</td>
<td>Short (1 word)</td>
<td>Brothers that push sisters with grape juice boxes are selfish.</td>
</tr>
<tr>
<td>Subject Relative</td>
<td>Long (5 words)</td>
<td>Brothers with grape juice boxes that push sisters are selfish.</td>
</tr>
<tr>
<td>Object Relative</td>
<td>Short (0 words)</td>
<td>Sisters with grape juice boxes that brothers push are selfish.</td>
</tr>
<tr>
<td>Object Relative</td>
<td>Long (4 words)</td>
<td>Sisters that brothers with grape juice boxes push are selfish.</td>
</tr>
</tbody>
</table>

**Time Compression:** The sound files for the sentences were time compressed to a compression ratio of 65/100 (~250 wpm) and 55/100 (~290 wpm) of the original playing time using the PRAAT software (v6.0.21, Institute of Phonetic Sciences, University of Amsterdam). All the sentences were also time compressed to a ratio of 100/100 (i.e. normal speech rate, 150 wpm) to ensure that any changes in the sound file due to the time compression process were consistent across all speech rates.

**Audibility Check**

The presentation level for the sound files was chosen to be 65 dB HL. To ensure audibility at this intensity level for all participants, an audibility check was conducted in a sound attenuated room before the start of the main experiment. The participants were presented successively with 15 short, 5 words-long sentences at 65 dB. These sentences had a simple Subject-Verb-Object construction (e.g. The whale ate the fish). For each compression ratio i.e. 100%, 65% and 55%, there were 5 sentences, that were played consecutively, blocked by speech rate. The order of presentation was kept the same for all participants such that they received the sentences at 100% compression first, followed by 65% and then 55%. After each sentence was
presented, the participant had to repeat the sentence aloud. Both the young and the older adults showed 100% accuracy in this task for sentences in all the speech rates.

**Procedure**

The main experiment was carried out in the same sound-attenuated room where the hearing screen and the audibility check were conducted. All the sound files were presented at 65 dB binaurally over Eartone 3A insert earphones (E-A-R Auditory Systems; Aero Company, Indianapolis, IN) via a Grason-Stadler GS-61 clinical audiometer (Grason Stadler Inc., Madison, WI). The participants were instructed to listen to and then to recall the sentence out loud to the best of their ability. Their responses were recorded in a Sony ICDPX333 Digital Voice Recorder.

Each participant was presented with a total of 120 sentences in 3 blocks corresponding to the three speech rates. The stimulus sentences were blocked by speech rate such that each block had 40 sentences at a given compression ratio with 10 sentences of each of the 4 syntactic structures – Subject Relative-Short Separation (SR-S), Subject Relative-Long Separation (SR-L), Object Relative-Short Separation (OR-S) and Object Relative-Long Separation (OR-L). Within each block, the order of sentence types was randomized. The stimuli in each block were preceded by 3 practice sentences to acclimate the participants to the compression ratio in each block. These sentences were 8 words long and had a simple Subject-Verb-Object construction (e.g. Boys represent girls near parks and are musical).

The sentences were counterbalanced across participants such that at the end of the experiment, each sentence was presented in Subject Relative-Short Separation, Subject Relative-Long Separation, Object Relative-Short Separation and Object Relative-Long Separation structures and in each of the 3 compression ratios at least once. To eliminate order effects, the
order of presentation speech-rates was also counterbalanced across participants. That is, this experiment had a ‘within-subjects’ design such that each participant received the same number of sentences in all conditions of syntactic structure and compression ratios and no participant received a given sentence in more than one condition.

Prior to the experiment, a written informed consent was obtained from all participants in accordance with a protocol approved by the Brandeis University Institutional Review Board.

**Scoring of the responses**

The responses recorded from the participants for the 120 trials were transcribed and scored. To score, the words in the transcribed responses for each trial were matched with the words in the actual sentence. A point of 1 was given for each correct word. A perfect score of 10 (for 10 correct words) was given if *all* the words of the stimulus sentence had been spoken. The order in which all the words were spoken was not taken into consideration for awarding a score of 10, such that a sentence that had a different syntax compared to the stimulus was also given a perfect score of 10 if all the words of the stimulus sentence were spoken. Extra words spoken were not penalized.
RESULTS

Accuracy of recall was measured as the percentage of total sentences in which all the correct words were spoken. The left panel of Figure 1 shows the recall accuracy of young adults and older adults for the sentences in Subject-Relative structure with a short and long separation for each of the three speech rates. The right panel shows these data for the Object-Relative sentences.

Figure 1: Recall Accuracy of young adults and older adults for the Subject Relative-Short Separation, Subject Relative-Long Separation, Object Relative-Short Separation & Object Relative-Long Separation sentence structures presented in the three speech rates – compression ratio of 100% (Normal Speech), 65% and 55%. Error bars are one standard error.
As was expected, both the young and older adults showed at or near ceiling performance for the simpler syntactic structure i.e. Subject Relative-short separation. Difference in the performance of young adults and older adults are seen for sentences with more complex structures i.e. SR-L, OR-S, OR-L.

I conducted an omnibus mixed design analysis of variance (ANOVA) on the recall accuracy data shown in Figure 1 that included effects of participant group (2: young adults vs. older adults), syntactic structure (2: Subject-Relative vs. Object-Relative), separation (2: short vs. long) and Speech Rate or Compression Ratio (3: 100%, 65%, 55%). All variables were within-subject variables except for participant group which was a between-subjects variable.

The ANOVA confirmed a significant main effect of syntax, \(F(1, 30) = 18.229, p < 0.001, \eta^2 = 0.378\) which is consistent with previous findings that the performance of older adults in recall is poorer with Object-Relative constructions. The effect of syntax was the same across the two groups as Syntax X Group interaction did not reach significance. Neither the main effect of agency separation (\(F(1, 30) = 4.092, p=0.052, \eta^2 = 0.120\)) nor the Separation X Syntax interaction (\(F(1, 30) = 2.553, p=0.121, \eta^2 = 0.078\)) was significant. The Separation X Group interaction did not show a significant effect either, reflecting similar effects of agency separation for both young and older adults alike.

There was a significant main effect of speech rate, \(F(2, 30) = 27.416, p<0.001, \eta^2 = 0.477\). The effect of speech rate varied across the two groups as was evidenced by a Speech Rate X Group interaction, \(F(2, 60) = 14.54, p<0.001, \eta^2 = 0.326\), confirming the findings from studies in the past that reported a reduced performance of older adults relative to young adults in the recall of time-compressed speech (Wingfield et al., 1985).
I also conducted a series of follow-up ANOVAs and pairwise comparisons to further dissect these effects. For the sentences with a Subject-Relative structure, there was a significant main effect of speech rate on the recall, $F(2, 60) = 10.349$, $p<0.001$, $\eta p^2 = 0.256$, while the effects of separation did not reach significance. This shows that Subject-Relative sentences, though in a simpler syntactic structure, become challenging as the speech rate is increased. There was also a significant Speech Rate X Group interaction, $F(2, 60) = 7.726$, $p=0.001$, $\eta p^2 = 0.205$, indicating that the challenge with higher speech rate for simpler syntactic structures is greater for older adults than for young adults.

For the young adults, a three-way ANOVA conducted on recall showed a significant main effect of syntax, $F(1, 15) = 34.921$, $p<0.001$, $\eta p^2 = 0.700$ and separation, $F(1, 15) = 14.300$, $p=0.002$, $\eta p^2 = 0.488$, as well as a significant Syntax X Separation interaction, $F(1, 15) = 4.613$, $p=0.048$, $\eta p^2 = 0.235$. Speech Rate, however, did not show a significant effect on the recall of sentences in any of the four syntactic structures in young adults, $F(1, 15) = 2.996$, $p=0.065$, $\eta p^2 = 0.166$, or Syntax X Speech Rate interaction, $F(2, 30) = 0.403$, $p=0.672$, $\eta p^2 = 0.026$. This indicates that the effect of compression for Subject Relative sentences was only on the older adults, not the young adults.

For the Object-Relative structure, outcomes of the ANOVA were similar to those for Subject Relative sentences with the effect of speech rate and the Speech Rate X Group interaction being significant, $[F(2, 60) = 25.193$, $p<0.001$, $\eta p^2 = 0.456$ and $F(2, 60) = 9.846$, $p<0.001$, $\eta p^2 = 0.247$ respectively]. The effect of separation was not significant for Object-Relative structure as well.

For the older adults, in a three-way ANOVA conducted on recall, syntax alone showed a significant effect, $F(1, 15) = 7.134$, $p=0.017$, $\eta p^2 = 0.322$, and so did speech rate alone, $F(2, 30)$
= 25.196, \( p<0.001 \), \( \eta^2 = 0.627 \), although the effect of separation or Syntax X Separation interaction was not significant \([ F(1, 15) = 1.418, p=0.252, \eta^2 = 0.086 \) \& \( F(1, 15) = 0.713, p=0.412, \eta^2 = 0.045 \) respectively\]. More interestingly, the effect of Syntax X Compression interaction was not significant for older adults, \( F(2, 30) = 0.492, p=0.616, \eta^2 = 0.032 \). This shows that while Object Relative sentences are more difficult than Subject Relative sentences, and a higher speech rate is more difficult than a lower speech rate, the increasing difficulty with higher speech rate is the same for both Subject Relative and Object Relative sentences.
DISCUSSION

Aging affects the memory for speech. My experiment shows the differential effects of speech rate and syntactic complexity on the memory for speech by young and older adults. Increase in speech rate did not affect the performance of the young adults in sentence recall. However, syntactic complexity (Subject Relative vs. Object Relative) or separation (short vs. long) affected the recall accuracy of the young adults. The fact that young adults have a better working memory and processing speed than older adults (Salthouse, 1994) may have allowed them to withstand the effects of time compression, at least for the compression ratios I employed in this experiment.

Recall accuracy of older adults was similar for both Subject-Relative and Object-Relative structures at the normal speech rate. In order to perform at the same level for the syntactically complex sentences, older adults may have required greater cognitive effort (e.g., Amichetti, Stanley, White, & Wingfield, 2013).

As the processing challenge was increased by increasing the speech rate, the recall accuracy of the older adults was impaired for sentences with both syntactic structures. This was especially so for the more complex Object-Relative structure. Processing of high speech rate would exceed the upper limit of total cognitive resources available for processing of speech signal. At this stage, the balance between the top-down and bottom-up processing would be disrupted, causing failures in storing the sentences in memory. When the speech rate was high, sentences that have the simpler more canonical word-order of a Subject-Relative structure, also
became difficult to comprehend. There was even greater difficulty in understanding sentences with the non-canonical, Object-Relative structure.

The recall accuracy of older adults was similarly affected by speech rate for both the Subject Relative and Object Relative sentences. This points to compensatory mechanisms, utilizing preserved linguistic knowledge, employed by older adults that would act similarly for both the Subject Relative and Object Relative sentences as the speech rate increases. Such a compensation is not seen in young adults, whose performance is unaffected by speech rate but suffers when the syntactic complexity is increased. This suggests that older adults, having a constant necessity for such compensation, have refined these neutralizing mechanisms with continuous use over the years.

In the wake of the challenges faced in comprehension of sentences at a high speech rate, there would be gaps in the sentences recalled by older adults. The older adults would, thus, be compelled to fill in these gaps using contextual cues and their preserved linguistic knowledge. They would, then, show a reconstructive recall of the sentences. It would be interesting to look at the errors or reconstructions made by the older adults in the task in this experiment to get an insight into how the older adults compensate for decline in cognitive abilities to comprehend speech successfully.

It is possible that they would make phonological substitutions to fill in the gaps with words that match the sound that they heard. For example, substituting ‘sleek leather’ with ‘sweet weather’ in the sentence: ‘Queens that kings with sleek leather gloves motivate are energetic’. Predominance of such reconstructions would point to a compensation for perceptual or sensory deficits using their preserved knowledge of the phonology of language.
The other possibility is that they would make semantic substitutions, filling in the gaps with words that are similar in meaning or belonging to the same class of words as the word that the older adults heard, but missed. An example of this would be the substitution of ‘nice’ with ‘good’ (words with the same meaning) in the sentence: ‘Boys that girls with full lunch boxes push are nice’ or the substitution of ‘narrow’ with ‘wide’ in the sentence: ‘Women that men in narrow office buildings interrupt are jovial’. Such errors, if predominant, would indicate that the older adults are comprehending the gist of the sentence despite the decline in their working memory and processing speed. It is quite likely that a mix of both kinds of substitutions, phonological and semantic, would occur simultaneously, such as the substitution of ‘anxious’ with ‘nervous’, both of which have a similar meaning and the same ‘ous’ sound in the end.

Perhaps, such a reconstruction is occurring even when the processing challenge is low, such as with simpler syntax or at normal speech rate. The performance seen in these scenarios may be due to a more accurate reconstruction that, in effect, produces a reproductive or veridical recall. The process of achieving this accuracy may be contributing to the increased cognitive effort expended by the older adults. Future studies could look at the difference in reconstructions made by the older adults and young adults to shed light on the distinct processes that could be at work for the comprehension of speech under difficult listening conditions in the two groups.
REFERENCES


