 ©2011 The Acoustical Society of Japan

ACOUSTICAL LETTER

Effects of pause duration and speech rate on sentence intelligibility in younger and older adult listeners

Akihiro Tanaka[1,2], Shuichi Sakamoto[1] and Yôiti Suzuki[1]

[1]*Research Institute of Electrical Communication, Tohoku University, Katahira 2–1–1, Aoba-ku, Sendai, 980–8577 Japan*
[2]*Waseda Institute for Advanced Study, Waseda University, Nishi-Waseda 1–6–1, Shinjuku-ku, Tokyo, 169–8050 Japan*

(*Received 18 May 2011, Accepted for publication 20 June 2011*)

Keywords: Speech rate, Pause duration, Sentence intelligibility
PACS number: 43.71.Gv, 43.71.Lz, 43.71.Sy [doi:10.1250/ast.32.264]

1. Introduction

Many older people have difficulty understanding spoken language in everyday communicative situations, especially when the speech is presented at a rapid rate. Time-compressed speech is less intelligible than normal rate speech especially for older adults (e.g., [1,2]), although this aging effect could be due to distortions by signal processing [3]. Age-related changes have been observed on various measures spanning perceptual to cognitive stages, including working memory [4] and attention [5]. Such cognitive aging might result from a general slowing of processing speed [6].

As opposed to speech compression, speech expansion may help in the comprehension of spoken language. Clear speech, a speaking style that many speakers adopt in order to be understood more easily in difficult communication situations [7], is usually slower [8] and more intelligible [7] than conversational speech. Although this line of evidence implies that speaking slowly is a good method in terms of intelligibility, previous studies have shown conflicting results on the effect of time expansion. Some studies have shown that time expansion improves intelligibility (e.g., [9]) while others have shown a decrease [10–12] or a null effect [13].

One reason for the conflicting results of time expansion would be that there are some distortions caused by signal processing at the time of speech expansion. This can explain the negative effect of time expansion at syllable and word levels. However, spoken language comprehension involves higher cognitive as well as earlier perceptual processes [14]. Thus, another reason for the conflicting results could be that the effects of time expansion might reflect performance on two distinct levels of processing: perceptual and cognitive levels. A single syllable benefits not from higher cognitive processing but from bottom-up perceptual processing. However, the benefit of time expansion at sentence or longer levels may reflect increased effectiveness in cognitive as well as perceptual processing. Because human information processing resources are limited [15], fewer resources are available for cognitive processes when the perceptual load is higher, leading to reduced efficiency and speed of cognitive processing [16]. In such cases, even when a syllable or a word has been correctly perceived, it may not be encoded in short-term memory to be available for higher processes. Rabbitt [17] showed that memory performance for spoken digits is worse when they were presented under a noisy but correctly perceivable condition than when they were presented without noise. McCoy *et al.* [18] showed that short-term memory performance in persons with hearing loss was lower than in those with better hearing even when perceptual performances in both groups were almost perfect. These results clearly show the interaction between perceptual and cognitive levels. McCoy *et al.* [18] concluded that the extra effort to achieve perceptual success may come at the cost of cognitive resources that might otherwise be available for encoding the speech content in memory.

In our study, we investigated the effect of pauses on speech-expansion benefit in spoken language comprehension. Pauses between phrases and speech rate were independently manipulated in the experiment. These manipulations of pauses and speech rate change the efficiency at both perceptual and cognitive levels. The manipulation of speech rate may mainly change the efficiency in perceptual processing. The manipulation of pauses may mainly change the efficiency in cognitive processing, as pauses can provide a time for cognitive processing [19], although the presence or absence of pauses also provides a perceptual cue for segmentation of the speech signal. Sentence intelligibility was compared between younger and older adult listeners to investigate age-related changes in the interaction between the effects of speech expansion and pause duration.

2. Methods

2.1. Participants

Eighteen listeners participated in the experiment. One group consisted of ten younger adults reporting normal hearing and ranging in age from 20 to 29 years (M (mean) = 23.0 years). A second group consisted of eight older adults with nearly normal hearing, ranging in age from 65 to 74 years (M = 68.9 years). The mean hearing level for frequencies within the speech range (500, 1,000, 2,000 Hz) in the second group was 14.1 dB (SD (standard deviation) = 4.9). All participants were native speakers of Japanese.

2.2. Materials

Five hundred and forty sentences were chosen from "The Phoneme-Balanced 1000 Sentence Speech Database" (NTT-

AT Co., Ltd.). The sentences were spoken in Japanese by a trained female speaker of Japanese. All speech materials were digitized with 16-bit resolution at a sampling rate of 16 kHz. The average speech rate in the original sentences was 7.0 morae per second. The A-weighted equivalent continuous sound pressure level was equated among the sentences excluding pauses.

In order to minimize the negative effects due to signal distortion, we used the STRAIGHT speech analysis-synthesis method [20], which had won first place among four synthetic vocal systems in the blind listening test conducted by RENCON'04 [21], to time expand the speech signals. Twenty-five types of stimuli were made for each of the 540 sentences. First, a sentence was cut into phrases (2.77 phrases on average, ranging between 2 and 4). The mean duration of phrases was 1.52 seconds ($SD = 0.67$). Subsequently, analysis and resynthesis were applied to change the duration of the phrases. The length of each modified phrase was 0, 100, 200, 300, or 400 ms longer than the original. Finally, pauses (i.e., silent portions) were inserted between the phrases. The length of pauses was 0, 100, 200, 300, or 400 ms. Thus, these 25 types of stimuli were made from each of the single sentences by a combination of the five expansion times and five pause lengths.

In addition to these synthesized stimuli, a pause control sentence and the original sentence were used as the control stimuli. Both these stimuli consisted of the original, unsynthesized phrases. The pauses between the phrases were not controlled in the original sentences, whereas these were equated to 400 ms in the pause control sentences. The mean duration of the pauses between the phrases was 201 ms in the original sentence stimuli. Examples of the speech signals used in some of the conditions are shown in Fig. 1.

Each of the 540 sentences was randomly assigned to each of the 27 experimental conditions. This sentence-to-condition assignment was randomized among participants.

2.3. Procedure

Listeners were tested individually in a sound-attenuating room. Auditory stimuli were played out via the TDT System III (Tucker-Davis Technologies, Gainesville, FL). The speech signals were presented at 75 dB (A-weighted equivalent continuous sound pressure level) to younger listeners. For older listeners, the speech signals were presented at their most comfortable levels because of the large individual differences in their hearing levels. As the result of an informal preliminary test revealed high intelligibility, background noise was added to prevent ceiling effects. The noise signal was filtered random noise simulating the long-term average of typical speech. The speech and noise signals were mixed electrically and presented through headphones (Sennheiser, HDA-200) over the left ear (younger listeners) or the better ear (older listeners).

Because the result of a preliminary test revealed high intelligibility (about 80 percent), a background noise was added to prevent the ceiling effect. The signal-to-noise ratio (SNR) was determined to avoid the ceiling effect and the floor effect and to match the performances of younger and older listeners. In particular, the SNR for younger listeners was −6 dB for half of the trials and −7 dB for the remaining half of the trials, on the basis of the results of another preliminary experiment. Since individual differences in speech intelligibility were larger in older listeners than in younger listeners, the SNR in older listeners was determined individually through practice trials. Using the results obtained under the four SNR conditions (−6, −3, 0, and +3 dB), the SNR was determined at individual optimal levels to avoid the ceiling and floor effects.

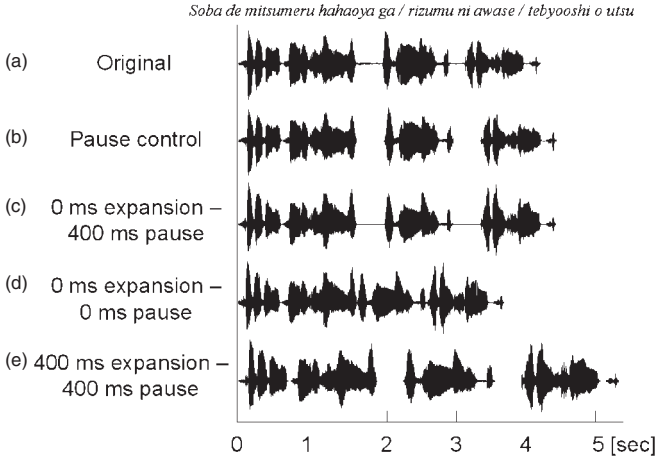

Fig. 1 Example of speech waveforms of (a) original sentence, (b) pause control sentence, (c) 0-ms-expansion and 400-ms-pause sentence, (d) 0-ms-expansion and 0-ms-pause sentence, and (e) a 400-ms-expansion and 400-ms-pause sentence.

In each of the main trials, a spoken sentence in noise was presented through headphones. Participants were asked to write down the sentence at their own pace. They were encouraged to guess at any words that were not clear or intelligible. The next trial began after the participants pressed a button.

The experiment consisted of 10 sessions. Each session took about 30–60 min depending on the pace of each participant. Participants completed two sessions per day. Each session consisted of 54 trials, comprising two trials for each of the 27 conditions. The order of sessions was counterbalanced. The order of trials within a session was randomized.

2.4. Data analysis

For each sentence, the percentage of correctly written keywords was calculated. The keywords were content words (including nouns, verbs, adjectives, and adverbs) ranging between four and seven words per sentence. Grammatical and spelling errors were ignored. Sentence intelligibility under each condition was derived for each listener based on the mean value through the 20 sentences under that condition.

3. Results

As a result of the manipulation of the SNR to match the intelligibility between groups, the mean SNR was −6.5 dB for younger listeners and −1.5 dB (ranging between −4.5 and +3.5 dB) for older listeners. As a result of this manipulation, the mean intelligibility was not significantly different between groups [$F(1, 16) = 0.27$, n.s.].

Figure 2 shows the mean intelligibility for younger and older listeners as a function of the duration of the pauses and the amount of expansion. A three-way analysis of variance

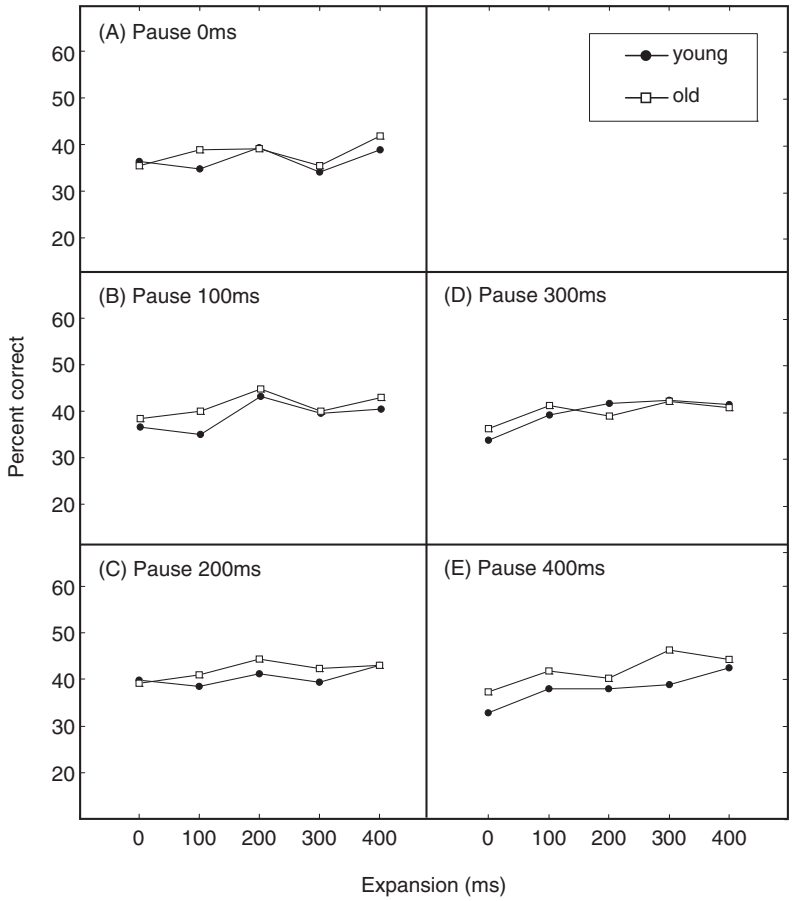

Fig. 2 Mean intelligibility (percentage of correctly written keywords) of younger and older listeners as a function of the duration of the pauses and the amount of expansion. Filled circles represent the results for younger listeners. Open squares represent those for older listeners.

revealed that the main effect of pauses was significant [$F(4, 64) = 5.17$, $p < 0.01$]. Multiple comparison (Fisher's LSD, $p < 0.05$) revealed that sentence intelligibility was better under conditions with pause (100, 200, 300, and 400 ms) than without pause (0 ms). The main effect of expansion was also significant [$F(4, 64) = 12.52$, $p < 0.01$]. Multiple comparison revealed that sentence intelligibility was better when the speech signal was expanded (100, 200, 300, and 400 ms) than when it was not (0 ms). Also, intelligibility at 200 and 400 ms expansions was significantly higher than at 100 ms expansion. The main effect of Group was not significant [$F(1, 16) = 0.32$, $p = 0.58$].

Importantly, the interaction between pause and expansion was significant [$F(16, 256) = 1.69$, $p < 0.05$]. Simple main effect analyses showed that the effect of expansion was significant at 0 ms [$F(4, 320) = 2.75$, $p < 0.05$], 100 ms [$F(4,320) = 4.21$, $p < 0.01$], 300 ms [$F(4,320) = 4.03$, $p < 0.01$], and 400 ms pauses [$F(4, 320) = 5.72$, $p < 0.01$]. Multiple comparison revealed that sentence intelligibility was higher at 400 ms than at 0 ms expansion when the pause was 0 ms, higher at 200 ms and 400 ms expansions when the pause was 100 ms, and higher at 100 ms and longer expansion when the pause was 300 ms and 400 ms. In sum, speech expansion as short as 100 ms yielded higher intelligibility than non expanded speech when the pause between phrases was long (i.e., 300 and 400 ms).

There was no significant difference between the pause control sentences and the original sentences in both younger [$t(9) = 0.31$, n.s.] and older [$t(7) = 0.75$, n.s.] listeners, confirming that keeping the pauses between phrases constant (400 ms) does not affect the intelligibility.

4. Discussion

We showed the interaction between pause duration and speech-expansion in spoken language comprehension. It is noteworthy that higher sentence intelligibility was obtained with a relatively short time expansion (i.e., 100 ms) when the pause between phrases was long enough (i.e., 300 and 400 ms) in both younger and older listeners. By inserting a pause between phrases, participants can use more time for higher cognitive processes. Consequently, more resources can be allocated to perceptual processing of the incoming speech signal. As processing resources were sufficiently allocated to the perceptual processing of time-expanded speech fragments when the pauses were long, facilitation effects of time expansion were salient, particularly under those conditions. This explanation is similar to the "effortfulness hypothesis" [18], which assumes an interaction between perceptual and cognitive processes.

One reason for the conflicting results of time expansion among studies is that there are some distortions caused by signal processing at the time of speech expansion. Our results

raise another possibility that does not exclude the distortion explanation. In previous studies that failed to demonstrate a facilitation effect of time expansion of the speech portion, cognitive resources might have been insufficient for higher cognitive processing of time-expanded, and thus easily perceivable, speech fragments. Whereas McCoy *et al.* [18] showed that perceptual difficulty affects cognitive processing, cognitive difficulty may also affect perceptual processing.

Our results have an implication on the design of an audiovisual speech rate conversion system. In audiovisual materials, if pauses (i.e., no-sound portions) between phrases were deleted in an amount equal to the duration of speech sound expansion, the onset of the next phrase becomes synchronous with the corresponding part of the (unchanged) visual signal. A speech rate conversion method, which manipulates speech signals following the above rule, has been developed for broadcasting [22]. Interestingly, our results showed that this pause-to-expansion transposition enhanced sentence intelligibility. Intelligibility in sentences with 200 ms pause and 200 ms expansion was higher than those with 400 ms pause and 0 ms expansion [$t(17) = 3.57$, $p < 0.01$]. It is noteworthy that lipread information as well as the benefit from speech expansion is available under this condition since the intelligibility is higher when the visual speech signal is presented along with the time-expanded auditory speech signal [23,24].

In summary, we examined the interaction between cognitive and perceptual processing levels. Results showed an interaction between pause duration and speech expansion. Speech expansion as short as 100 ms was more effective than non-expanded speech when the pause between phrases was 300 or 400 ms. This tendency was common between age groups as long as the difficulty was matched.

Acknowledgements

This work was supported by Grants-in-Aid for Specially Promoted Research No. 19001004 from MEXT Japan to SY, and the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University (H22-A09) to TA. The authors thank Dr. H. Kawahara for permission to use the STRAIGHT vocoding method. The authors would also like to thank Atsushi Imai and Tohru Takagi at the NHK Science and Technical Research Laboratories for their helpful comments on our research.

References

[1] D. F. Konkle, D. S. Beasley and F. H. Bess, "Intelligibility of time-altered speech in relation to chronological aging," *J. Speech Hear. Res.*, **20**, 108–115 (1977).

[2] A. Wingfield, L. W. Poon, L. Lombardi and D. Lowe, "Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time," *J. Gerontol.*, **40**, 579–585 (1985).

[3] B. A. Schneider, M. Daneman and D. R. Murphy, "Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?" *Psychol. Aging*, **20**, 261–271 (2005).

[4] L. Hasher and R. T. Zacks, "Working memory, comprehension, and aging: A review and a new view," *Psychol. Learn. Motiv.: Adv. Res. Theory*, **22**, 193–225 (1988).

[5] J. Cerella, "Generalized slowing and Brinley plots," *J. Gerontol. Psychol. Sci.*, **49**, 65–71 (1994).

[6] T. A. Salthouse, *Theoretical Perspectives on Cognitive Aging* (Erlbaum, Hillsdale, NJ, 1991).

[7] J. C. Krause and L. D. Braida, "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," *J. Acoust. Soc. Am.*, **112**, 2165–2172 (2002).

[8] M. A. Picheny, N. I. Durlach and L. D. Braida, "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Hear. Res.*, **29**, 434–446 (1986).

[9] J. F. Schmitt, "The effects of time compression and time expansion on passage comprehension by elderly listeners," *J. Speech Hear. Res.*, **26**, 373–377 (1983).

[10] Y. Nejime and B. C. Moore, "Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss," *J. Acoust. Soc. Am.*, **103**, 572–576 (1998).

[11] M. Picheny, N. Durlach and L. Braida, "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech," *J. Speech Hear. Res.*, **32**, 600–603 (1989).

[12] S. Kemper and T. Harden, "Experimentally disentangling what's beneficial about elder-speak from what's not," *Psychol. Aging*, **14**, 656–670 (1999).

[13] J. A. Small, S. Kemper and K. Lyons, "Sentence comprehension in Alzheimer's disease: Effects of grammatical complexity, speech rate and repetition," *Psychol. Aging*, **12**, 3–11 (1997).

[14] M. K. Pichora-Fuller, "Cognitive aging and auditory information processing," *Int. J. Audiol.*, **42** (Suppl. 2), S26–S32 (2003).

[15] F. I. M. Craik, "The role of cognition in age-related hearing loss," *J. Am. Acad. Audiol.*, **18**, 539–547 (2007).

[16] M. K. Pichora-Fuller, B. A. Schneider and M. Daneman, "How young and old adults listen to and remember speech in noise," *J. Acoust. Soc. Am.*, **97**, 593–607 (1995).

[17] P. M. A. Rabbitt, "Channel capacity, intelligibility and immediate memory," *Q. J. Exp. Psychol.*, **20**, 241–248 (1968).

[18] S. L. McCoy, P. A. Tun, L. C. Cox, M. Colangelo, R. A. Stewart and A. Wingfield, "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," *Q. J. Exp. Psychol.*, **58A**, 22–33 (2005).

[19] A. Wingfield, P. A. Tun, C. K. Koh and M. J. Rosen, "Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech," *Psychol. Aging*, **14**, 380–389 (1999).

[20] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency," *Speech Commun.*, **27**, 187–207 (1999).

[21] K. Noike, R. Hiraga, M. Hashida, K. Hirata and H. Katayose, "A report of NIME04 Rencon: Listening contest to evaluate performance rendering systems," *Proc. 19th Annu. Conf. Japanese Society for Artifical Intelligence*, 2B3-07 (2005).

[22] A. Imai, R. Ikezawa, N. Seiyama, A. Nakamura, T. Takagi, E. Miyasaka and K. Nakabayashi, "An adaptive speech rate conversion method for news programs without accumulating time delay," *J. IEICE*, **83-A**, 935–945 (2000).

[23] S. Sakamoto, A. Tanaka, K. Tsumura and Y. Suzuki, "Effect of speed difference between time-expanded speech and moving image of talker's face on word intelligibility," *J. Multimodal User Interfaces*, **2**, 199–203 (2008).

[24] A. Tanaka, S. Sakamoto, K. Tsumura and Y. Suzuki, "Visual speech improves the intelligibility of time-expanded auditory speech," *Neuroreport*, **20**, 473–477 (2009).