Rapid perceptual learning of time-compressed speech and the perception of natural fast speech in older adults with presbycusis

TALI ROTMAN[1,*], LIMOR LAVIE[1] AND KAREN BANAI[1]

[1]*Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel*

Older people, especially ones with age-related hearing loss (ARHL), often have difficulties understanding naturally fast speech (NFS). This difficulty has been attributed to both sensory and cognitive factors. We now ask if rapid perceptual learning, assessed with a time-compressed speech (TCS) task, also contributes to the perception of NFS in older adults with ARHL, while accounting for the potential contribution of other cognitive factors. 45 participants with and without experience with hearing-aids completed the study. Significant rapid perceptual learning of TCS occurred within the first 20 sentences. This learning was significantly and positively correlated with NFS perception. Additionally, the perception of NFS was positively associated with vocabulary and memory span and negatively correlated with hearing thresholds. We found no significant differences between experienced-users of hearing-aids and non-users. Findings suggest that declines in rapid perceptual learning may play a role in the perception of NFS in people with ARHL, which is additional to the contribution of other cognitive variables.

INTRODUCTION

Perceiving rapid speech is a major challenge for older listeners (Gordon-Salant, 2001; Schneider *et al*. 2005; Janse, 2009). Many studies evaluated the effect of aging on the perception of rapid speech, in most cases using time-compressed speech (TCS). They attributed the perceptual difficulties of older listeners to both auditory (peripheral and central) and cognitive declines (Gordon-Salant, 2001; Wingfield & Tun; 2001; Janse, 2009; Gordon-Salant & Friedman, 2011).

Another factor that influences the recognition of rapid speech, even among older adults, is short-term perceptual learning (Peelle & Wingfield, 2005; Golomb et al., 2007). Manheim *et al*. (2018) showed that rapid perceptual learning is positively correlated with the perception of natural fast speech (NFS) in young normal hearing adults and in older adults, with either normal hearing or mild to moderate hearing loss. However, to our knowledge, the contribution of rapid perceptual learning to the perception of NFS has not been evaluated in older adults with more severe hearing impairments.

*Corresponding author: trotma01@campus.haifa.ac.il

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

Thus, the main goal of the current study was to evaluate the contribution of short-term perceptual learning of TCS to the perception of NFS in older adults with significant hearing impairments. Because previous studies showed that sensory (hearing acuity, speech audibility) and cognitive factors also accounted for variance in speech recognition in elderly listeners (Humes, 2002; Lunner et al., 2009; Humes & Dubno, 2010), we also evaluated the contribution of hearing thresholds, the initial recognition of TCS and cognitive factors (working memory, selective attention, verbal and non-verbal intelligence) to the perception of NFS.We hypothesized that rapid perceptual learning would contribute to the perception of NFS, even after considering the contribution of these other factors.

We also compared the rapid perceptual learning of TCS and perception of NFS between experienced hearing-aid users and non-users (participants who have never used hearing-aids and weren't fitted with hearing-aids during the study period). Despite the fact that all participants were tested through headphones, we hypothesized that long-term experience with hearing-aids may have an impact on perceptual learning or speech perception under more strained conditions such as NFS.

MATERIALS AND METHODS

Participants

Forty-five older adults (24 males, 21 females) aged 65-89 years (mean age: 75, SD: 7) with age-related hearing loss participated in this study. The mean pure tone threshold (at 0.5, 1, 2 & 4 kHz) in both ears was 50dB HL (range: 30-70dB HL, SD: 8) and the mean score in the word recognition test (WRT) in quiet was 90% (SD: 7). All participants had an excellent control of Hebrew, with no history of neurological problems (according to self-report) and achieved a score of 24 and above in the Hebrew Mini-Mental State Examination (MMSE) (Folstein *et al*., 1975). The mean years of education was 14 (range: 8-22, SD: 3).

Twenty-three of the participants had a long-term experience (1-12 years, mean: 4.41, SD: 3.15) with digital hearing-aids in both ears, while the rest of the participants had no experience with amplification devices whatsoever. Both groups were balanced in mean age, gender, mean years of education and mean score in the MMSE.

Perception of NFS and TCS

40 simple Hebrew sentences (taken from Prior & Bentin, 2006) were used. All sentences were five to six words long and had a common subject-verb-object grammatical structure. 20 sentences were semantically plausible (e.g., "The wide pipe fills a round bath"), while the other 20 were semantically implausible (e.g., "The sharp knife hurt the inflated ball"). All sentences were recorded in a sound attenuating booth and their root mean square (RMS) levels were normalized after recording using Audacity audio software. 20 sentences (10 semantically plausible and 10 semantically implausible) were recorded in natural fast speech by a female native Hebrew speaker at a mean rate of 221 words/minute (SD=23). The other 20 sentences were recorded

by another female native Hebrew speaker at a normal speech rate of 113 words/minute (SD: 12). The latter were timed-compressed to 50% of their original length in Matlab, using a WSOLA algorithm (Verhelst & Roelands, 1993). We compressed the normal speech by 50% in order to match it to the mean speech rate of the NFS.

All stimuli were presented bilaterally via headphones (hearing-aids were not used during stimuli presentation) at the most comfortable level (MCL) of each listener (range: 107-115dB sound pressure level (SPL) , mean: 112, SD: 1.4). The individual MCLs were determined in reference to the levels in dB SPL that were measured in the headphones output for a pure tone of 1kHz played from a personal laptop. An increment of 5% on the volume scale in the computer was equivalent to a change of 1dB SPL. First, the participants listened to the 20 sentences in NFS, and then to the 20 time-compressed sentences. Plausible and implausible sentences were presented in a random order (the same order across all participants). The listeners were asked to repeat what they heard after each single trial and the experimenter transcribed their replies. Performance with NFS was quantified as the percentage of words correctly identified across all sentences.

The perception of TCS was used to evaluate the rapid perceptual learning. We followed previous studies (e.g. Dupoux & Green, 1997; Peelle & Wingfield, 2005) in defining the perceptual learning effect as the difference between the percentage of correctly identified words in the final and first five sentences.

In order to determine if rapid learning of TCS accounted for individual differences in NFS, we used a two-stage linear regression model. In the first stage, we created a base model with all potential explanatory variables, except for rapid learning of TCS. In the second stage, we added the rapid-learning slope.

Cognitive measures

Flanker task

A computerized version of the Flanker task (Eriksen & Eriksen, 1974) was used as a measure of selective attention. We created this computerized version in SuperLab software, according to the parameters presented by Scharenborg *et al.* (2015). Participants were asked to respond to the direction of a central arrow flanked by arrows pointing at the same direction (congruent trials), the opposite direction (incongruent) or = signs (neutral trials). The "flanker cost" for each participant was calculated as the mean log reaction time (logRT) in ms of the correct responses in the incongruent trials divided by the mean logRT of the correct responses in the neutral trails.

Subtests from the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III) in Hebrew:

1) *Digit span forward and backward* – This subtest was used to measure auditory working memory capacity. Participants listened (with their personal hearing-aids or with a personal amplifier) to a sequence of digits and were asked to recall the sequence correctly, with increasingly longer sequences being tested in each item.

Each item included two trials. In digit span forward the participants were asked to recall the sequence in the same order of presentation, while in the backward task they were asked to recall it in reverse order.

2) *Vocabulary* – This subtest evaluated the patient's semantic knowledge and verbal concept formation in Hebrew. The participants were asked to give oral definitions of up-to 33 words. Each word was presented both visually (written on a card) and orally (by the experimenter).
3) *Block Design* – The participants were requested to copy a two-colors pattern using colored blocks. The patterns became more and more complex and were time-limited (0.5/1/2 minutes according to the difficulty level of the specific item). The score in each item reflected the accuracy and speed of performance. This subtest assessed the participant's ability to understand complex visual information.

All three subtests were administrated and scored according to the test manual; raw scores were converted to standard scores according to the participant's age group.

RESULTS

Cognitive measures

Flanker task

The mean accuracy of the responses pooled over all participants was 94% (SD: 12). Accuracy was lowest and most variable in the incongruent condition (87%, SD: 24), while accuracy was best and least variable in the congruent condition (98%, SD: 8); accuracy for the neutral condition was close to that of the congruent condition (97%, SD: 9). The mean Flanker cost was 1.01 (SD: 0.007). Higher Flanker cost means poorer selective attention.

Subtests from the WAIS-III

The mean standard scores and standard deviations in digit span, vocabulary and block design subtests across all participants were: 9.17 (SD: 2.4), 9.17 (SD: 2.7) and 9.31 (SD: 2.4), respectively.

Perception of NFS and rapid perceptual learning of TCS

The mean percentage of the correctly identified words across all 20 sentences in NFS was 36% (SD: 18, median: 34%, inter-quartile range (25th-75th): 27%).

For each participant, we calculated the rapid perceptual learning of TCS as the difference between the mean percentage of correctly identified words across the first five sentences to that of the final five sentences. This difference was defined as the rapid-learning delta (Fig. 1A). The mean rapid-learning delta across all participants was 22% (SD: 16). Paired-samples *t*-test showed that the rapid-learning delta was statistically significant (t (88) = 4.31, p<.0001) with a large effect size (Cohen's d=1.37).

We also calculated the slope of the learning curve over the 20 TCS sentences as follows: mean recognition accuracy was calculated for each mini block of 5 sentences, and the slope of this curve was computed (Fig. 1B). The mean rapid-learning slope across all participants was 7% (SD: 5), equivalent to an improvement rate of 1-2 words/sentence.

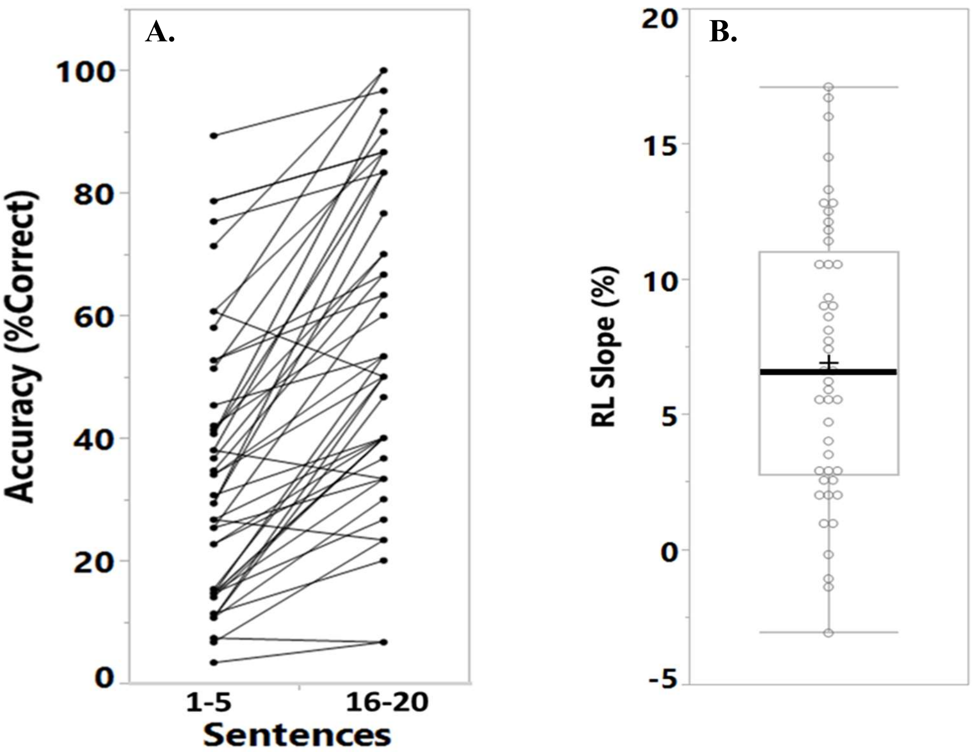

Fig. 1. A. Recognition of time-compressed speech. Line show mean individual performance during the first and the final five-sentences mini blocks. **B. Rapid-learning slope of time-compressed speech.** The box represents the interquartile range (25th-75th percentile). Whiskers are 1.5 the interquartile range. The thick line within the box marks the median and the + sign marks the mean. Each dot represents one participant (45 in total).

No significant differences were found between experienced and non-users in either rapid perceptual learning of TCS nor in perception of NFS ($t < 1$).

Individual differences in NFS perception as a function of hearing, cognition and learning

Recognition of NFS and the recognition accuracy in the first 2 time-compressed sentences were highly correlated ($r = 0.826$, $p < .001$). Likewise, the recognition of

NFS was also significantly or marginally correlated with age (r = 0.272, p=0.071), education (r = 0.399, p=0.007), hearing (PTA) (r = -0.499, p< .001), digit span (r = 0.444, p=0.002) and vocabulary (r = 0.465, p=0.001). The final linear regression model accounted for 85% of the variance in NFS perception, adjusted for the large number of predictors used relative to the sample size (F (7,37) = 37.88, p <.001). Of this, 7% were a unique contribution of rapid learning (F (1,37) = 20.99, p <.001). The full model is shown in Table 1. TCS perception at baseline was the strongest predictor, followed by rapid learning, hearing and vocabulary.

Stage	**Predictor**	**R^2**	**Adjusted R^2**	**F**	**β**	**t**
1	age				-0.08	-1.27
	education				0.14	1.93^m
	hearing				***-0.21***	***-3.66******
	digit span				0.04	0.58
	vocabulary				***0.20***	***2.81*****
	TCS (baseline)				***0.62***	***8.7******
2	**rapid learning**				***0.27***	***4.58******
Full model		**0.88**	**0.85**	**37.88***		

Table 1: Summary of the regression model for natural fast speech perception. R^2 – unadjusted proportion of variance; β – standardized regression. Coefficients; ^{m}p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001.

DISCUSSION

Consistent with our initial hypothesis, rapid perceptual learning of TCS and the perception of NFS were positively associated in older adults with significant age-related hearing loss. This extends the outcomes of two previous studies: Karawani *et al.* (2017) showed that older participants who improved less over the course of short-term speech-in-noise (SIN) training had poorer starting performance on the trained task as well as poorer performance on untrained speech in noise (SIN) tasks. Manheim *et al.* (2018) showed the same positive correlation between rapid perceptual learning of TCS and NFS perception in young adults with normal hearing, as well as in older adults with normal and impaired hearing. The results of the current study extend the results of Manheim *et al.*'s study to older adults with more severe hearing loss and account for the potential contributions of additional factors that have not been considered before in the context of perceptual learning for speech. Although correlations do not indicate causation, such correlations should exist if perceptual learning is to support the perception of degraded speech, as suggested by Samuel & Kraljic (2009).

Study participants completed tasks which evaluated the level of vocabulary, working memory, selective attention and non-verbal intelligence. We found that vocabulary and digit spans were positively correlated with the perception of NFS. Vocabulary was also a significant predictor of individual differences in NFS recognition. These findings are consistent with the 'ease of language learning' model (Rönnberg, 2003; Rönnberg *et al*., 2008; Rönnberg & Jerker, 2013). According to this model, perception of speech in adverse conditions (e.g. background noise, accented speech, rapid speech) involves explicit processing in which cognitive resources, such as working memory and semantic knowledge, have a critical role. Significant correlation between working memory capacity and recognition of SIN among older adults with age-related hearing loss was demonstrated by Nagaraj (2017).
To our knowledge, the present study is the first to evaluate the effect of long-term experience with hearing-aids on the perception of NFS and rapid perceptual learning of TCS. We found no differences between experienced hearing-aid users and the non-users. This result may stem from the fact that both groups listened to the stimuli through headphones at their MCLs, and hearing aids were not used during the perception and learning tests. Further research is needed to examine rapid speech perception through hearing aids, both in experienced users and in first-time hearing-aid users.

In conclusion, the results in this study support the idea that the difficulty of older adults to perceive rapid speech is connected to perceptual learning, but further work is required to elucidate the full interplay of sensory (e.g. temporal and spectral processes), cognitive and learning in the perception of challenging speech.

ACKNOWLEDGEMENTS

The Israel Science Foundation (grant 206/18) and Mr. Omri Gavish supported this study.

REFERENCES

Dupoux, E., Green K. (**1997**). "Perceptual adjustment to highly compressed speech: effect of talker and rate changes". J. Exp. Psychol. Hum. Percept. Perform., **23**, 914.

Eriksen, B. A., and Eriksen, C. W. (**1974**). "Effects of noise letters upon the identification of a target letter in a nonsearch task," Atten. Percept. Psycho., **16**, 143-149.

Folstein, M., Folstein, S., and McHugh, P. (**1975**). "Mini-mental state. A practical method for grading the cognitive state of patients for the clinician," J. Psych. Res., 12, 189-198.

Golomb, J. D., Peelle, J. E., and Wingfield, A. (**2007**). "Effects of stimulus variability and adult aging on adaptation to time-compressed speech," J. Acoust. Soc. Am., **121**, 1701-1708.

Gordon-Salant, S. (**2001**). "Sources of age-related recognition difficulty for time-compressed speech," J. Speech. Lang. Hear. Res., **44**, 709-719.

Gordon-Salant, S., and Friedman, S. A. (**2011**). "Recognition of rapid speech by blind and sighted older adults," J. Speech. Lang. Hear. Res., **54**, 622-631.
Humes, L. E. (**2002**). "Factors underlying the speech recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am., **112**, 1112-1132.
Humes, L. E., and Dubno, J. R. (**2010**). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, Edited by S. Gordon-Salant et al. (Springer, New York, NY), pp. 211-257.
Janse, E. (**2009**). "Processing of fast speech by elderly listeners," J. Acoust. Soc. Am., **125**, 2361-2373.
Karawani, H., Lavie, L., and Banai, K. (**2017**). "Short-term auditory learning in older and younger adults," Proc. ISAAR, **6**, 1-8.
Lunner, T., Rudner, M., and Rönnberg, J. (**2009**). "Cognition and hearing-aids," Scan. J. Psycho, **50**, 395-403.
Manheim, M., Lavie, L., and Banai, K. (**2018**). "Age, hearing, and perceptual learning of rapid speech," Trends Hear., **22**, 1-18.
Nagaraj, N. K. (**2017**). "Working memory and speech comprehension in older adults with hearing impairment," J. Speech. Lang. Hear. Res., **60**, 2949-2964.
Peelle, J. E., and Wingfield, A. (**2005**). "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech," J. Exp. Psychol. Hum. Percept. Perform., **31**, 1315.
Prior, A., and Bentin, S. (**2006**). "Differential integration efforts of mandatory and optional sentence constituents," Psychophys., **43**, 440-449.
Rönnberg, J. (**2003**). "Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model," Int. J. Audiol., **42**, S68-S76.
Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (**2008**). " Cognition counts: a working memory system for ease of language understanding (ELU)," Int. J. Audiol., **47**, S99-S105.
Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., …, and Rudner, M. (**2013**). "The ease of language understanding (ELU) model theoretical, empirical and clinical advances," Front. Syst. Neurosci., **7**, 31.
Samuel, A. G., and Kraljic, T. (**2009**). "Perceptual learning for speech," Atten. Percept. Psycho., **71**, 1207-1218.
Scharenborg, O., Weber, A., and Janse, E. (**2015**). "The role of attentional abilities in lexically guided perceptual learning by older listeners," Atten. Percept. Psycho., **77**, 493-507.
Schneider, B. A., Daneman, M., and Murphy, D. R. (**2005**). "Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing?," Psychol. Aging., **20**, 261.
Verhelst, W. and Roelands, M. (**1993**). "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, **2**, 554-557.
Wingfield, A., and Tun, P. A. (**2001**). "Spoken language comprehension in older adults: Interactions between sensory and cognitive change in normal aging," Semin. Hear., **22**, 287-302.