Perceptual learning and speech perception: A new hypothesis

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

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

Perceptual learning for speech remains substantial even in older adults, but the functional significance of this observation is not well understood. It has been suggested that perceptual learning might serve to support listening in adverse conditions by promoting behavioural and neural plasticity, but this hypothesis is not consistent with the acoustic specificity of learning. Instead, we now suggest that in the context of speech perception, perceptual learning might be best viewed as one of the capacities that, like working memory, support speech perception in an on-line fashion. Consistent with this hypothesis, we present data that rapid perceptual learning of one speech task accounts for substantial individual differences in other speech tasks even after accounting for the potential correlations between different indices of speech perception.

INTRODUCTION

Perceptual learning, defined as relatively long lasting experience-dependent changes in the processing of sensory stimuli, has been documented across sensory modalities and age groups (Green *et al.,* 2019). Speech, the focus of this paper, is also subject to perceptual learning that can occur rapidly, even following few minutes of exposure (for review see Samuel and Kraljic, 2009). Over the last decades, attempts were made to develop perceptual training regimens for hearing rehabilitation (e.g., Sweetow and Sabes, 2006), but a more recent systematic review (Henshaw and Ferguson, 2013) concluded that the evidence for the efficacy of such programs is weak. One of the reasons cited is that while robust, perceptual learning is also quite specific. For example, although older adults with and without hearing loss retain substantial learning of both speech in noise (Burk and Humes, 2008; Karawani *et al.,* 2016) and time-compressed speech (Manheim *et al.,* 2018), transfer of learning is limited by the acoustic and semantic similarity of new materials to those experienced in training. These findings of robustness and specificity raise the question of the role of perceptual learning in speech recognition. If past learning fails to modify future speech recognition to a substantial degree, what (if any) is the functional role of perceptual learning?

While struggling with the implications of the specificity of training-induced learning, we observed strong correlations between speech perception and perceptual learning across age and hearing levels. First, although the transfer of learning on time-compressed speech is limited, rapid learning (following < 5 minutes of listening) of

---

*Corresponding author: kbanai@research.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.

time-compressed speech is highly correlated with the recognition of natural-fast speech (Manheim *et al.,* 2018). Second, the amount of within-session learning on one speech in noise task, was strongly correlated with naïve performance on another speech in noise task (Karawani *et al.,* 2017). It could have been suggested that the correlations we observed reflect either rapid learning on the tasks which we used to measure ‘perception’, or the transfer of learning during the rapid learning phase to the other, different, task. Both explanations seem unlikely because, in the two studies we refer to, the perceptual tests were too brief to elicit rapid learning on either speech in noise or natural-fast speech. Furthermore, the learning phase in these two studies seems too brief to result in transfer (Adank and Janse, 2009; Wright *et al.,* 2010; Banai and Lavner, 2014). We therefore proposed an alternative view, that rapid perceptual learning, which underlies perceptual adjustment to challenging or unusual speech (e.g., Mattys *et al.,* 2012), might serve as one of the factors contributing to speech perception in adverse conditions. By this account, perceptual learning might play a role in speech perception, similar to that played by cognitive functions such as attention and working memory (Ronnberg *et al.,* 2019). The study described here is an attempt to test one of the predictions of this account. Namely, if rapid learning is some form of general capacity, the correlations observed in our previous studies should be replicated under conditions with less similarity between the stimuli used to assess rapid learning and those used to estimate speech perception. To this end, we used a time-compressed speech (TCS) task to estimate rapid perceptual learning, and speech in noise (SIN) and naturally fast speech (NFS) as indices of speech perception under adverse conditions. TCS was selected as the learning task due to the large number of studies that documented rapid, robust and long lasting learning with this task across age groups (Altmann and Young, 1993; Dupoux and Green, 1997; Peelle and Wingfield, 2005).

METHODS

Participants

Seventy-eight young adults, all native Hebrew speakers (ages 18-35, M = 26, SD = 4; 38 female) participated in this study as unpaid volunteers. By self-report, all participants had normal hearing and no history of learning, language or neurological disorder. Most participants were undergraduate students at the University of Haifa and other academic institutes in the region or recent graduates.

Procedure

Each participant completed two sessions, held 5-9 days apart. On the first session, we assessed the perception of time-compressed speech. On the second session, the recognition of time-compressed speech was tested again in order to calculate a between-session learning index. Subsequently, each participant completed a speech in noise and a natural-fast speech test. The order of these tests was counterbalanced across participants. Stimuli were presented diotically through headphones (Sennheiser HD-205 or 215) in a quiet room on campus or at their homes. All aspects of the study were approved by the ethics committee of the Faculty of Social Welfare and Health Sciences, University of Haifa (IRB 199/12).

Tasks and stimuli

For all tasks, we used 5-6 word sentences in Hebrew (Prior and Bentin, 2006), produced by two native female Hebrew speakers, recorded and amplitude normalized using the Audacity software. Each sentence was presented only once throughout the two study sessions. During each task participants were instructed to transcribe each sentence after it was played. The percentage of correct words was computed and used in data analysis. Only perfectly reported words were counted as correct (for further details on the word-scoring method see Manheim *et al.*, 2018).

Rapid perceptual learning of time-compressed speech (TCS).

On each session, 10 different sentences were presented. We defined rapid learning as the difference in transcription accuracy between the two sessions. Stimuli for this task were recorded by talker 1 at an average natural speech rate of 111 words/minute (SD = 17) and then compressed to 30% of their natural duration using a WSOLA algorithm (Verhelst and Roelands, 1993).

Natural-fast speech perception.

Twenty sentences were presented by talker 2 at an average natural fast rate of 214 words/minute (SD = 26).

Speech-in-noise perception.

Twenty sentences were presented diotically by talker 1 mixed in 4-talker babble noise (for details see Karawani *et al.*, 2016). Speech and noise were presented simultaneously, in each participant's most comfortable level. Signal-to-noise ratio was -6 dB.

RESULTS

Rapid perceptual learning of time-compressed speech

As shown in Fig. 1, 75% of participants improved their performance between the two sessions with a median improvement of 10% (IQR = 1-20%; $Z = 6.05$, $p < 0.001$). These data are consistent with previous findings on the rapid learning of time-compressed speech and its retention over time, and suggest that there are substantial individual differences in the magnitude or rate of rapid learning across participants.

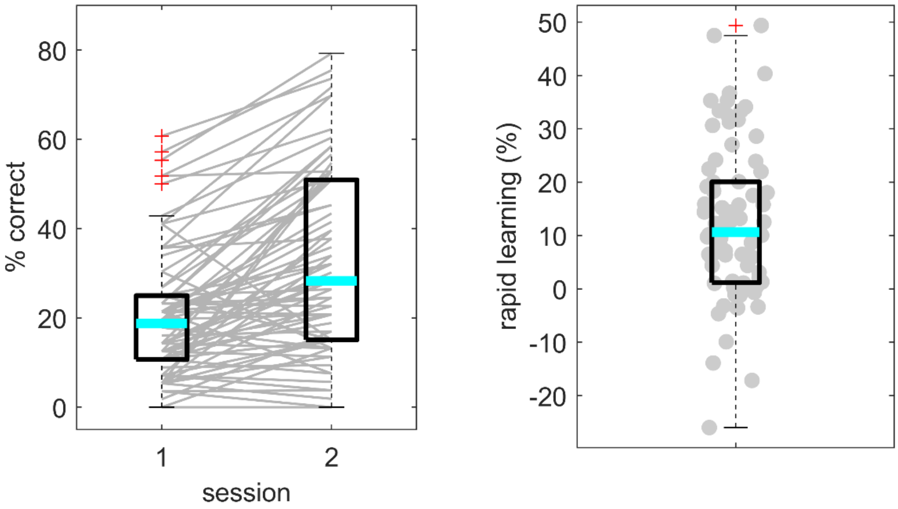

Fig. 1: Recognition (left) and rapid across-session perceptual learning (right) of time compressed speech. Box edges mark the inter-quartile range; thick line within each box marks the median; Whiskers are 1.5 the IQR. Grey symbols/lines show individual data.

Rapid learning and individual differences in speech recognition

Rapid learning and speech recognition were significantly correlated (SIN: $r = 0.35$, $p = 0.002$; NFS: $r = 0.44$, $p < 0.001$). These correlations (especially with NFS) may have been expected because rapid learning was also assessed using a speech task. Therefore, we attempted to statistically partial out the contribution of the TCS to these correlations. To this end, linear regression models were used in which baseline recognition of TCS (the first 5 sentences of the first session) were entered to the model first, and the rapid learning index was entered on a second stage. Although this is not a perfect control, baseline performance was not correlated with the rapid learning index ($r = 0.08$, $p = 0.48$). Details of these models, which generally conformed with the assumptions of liner models (tolerance > 0.95, VIF < 1.5), are shown in Table 1. The unique contribution of rapid learning to the perception of each type of perceptually difficult speech is depicted on Fig 2.

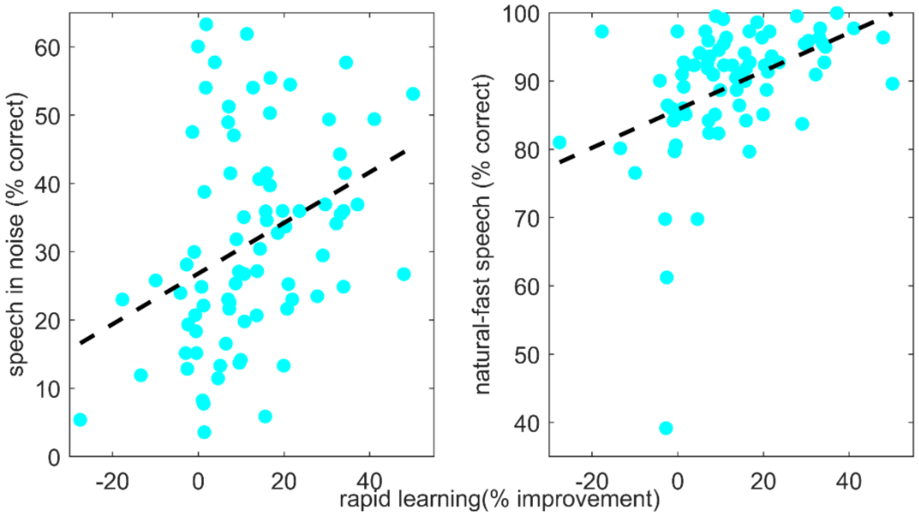

Fig. 2. Speech perception as a function of rapid learning of time-compressed speech. Partial correlation after controlling for baseline recognition of time-compressed speech is shown (for details of the formulas used see Manheim et al., (2018)).

	Predictor	**R²**	**F**	**B**	**t**
SIN	TCS perception			0.30	2.93**
	Rapid learning	0.11	10.22**	0.33	3.20**
	Full model	0.21	10.23***		
NFS	TCS perception			.16	1.58
	Rapid learning	0.18	17.61***	.43	4.20***
	Full model	0.20	10.66***		

Table 1. Speech perception as a function of baseline speech perception (TCS) and rapid perceptual learning. R^2 - the unique contribution of rapid perceptual learning followed the proportion of variance explained by the full model; β – standardized regression coefficients. $^{**}p < 0.005$, $^{***}p < 0.001$.

DISCUSSION

In the current study, we tried to broaden our understanding of the role of perceptual learning in speech perception under adverse conditions. Previous observations led us to suggest that age-related declines in rapid learning may contribute to speech perception deficits in older adults (Manheim *et al.*, 2018). This reflects a more general view of rapid perceptual learning as one of the factors contributing to individual differences in speech perception under adverse conditions. By this account, under a broad array of adverse conditions, rapid learning allows listeners to quickly calibrate speech processing based on the acoustic demands of the ongoing situation. The data presented in this manuscript is consistent with this idea. Although correlational in nature, it nevertheless shows that individual differences in rapid learning of time-compressed speech are associated with individual differences in the perception of natural-fast speech and speech in noise. The correlation for speech in noise is especially telling because time-compressed speech and speech in noise represent different domains of speech perception (as opposed to natural-fast and time-compressed speech which represent two forms of rapid speech) and possibly rely on different sensory processes and top-down strategies. Furthermore, the use of TCS as baseline in the statistical model should have accounted for the contribution of the processes shared by TCS and SIN.

More studies are required to test our hypothesis further. Going forward, it will be important to control for additional variables that are thought to be involved in both speech perception and perceptual learning (e.g., working memory and inhibition), as well as to account for the relationships between rapid learning and longer-term learning. It will also be of interest to consider populations with more variance in age and hearing levels than tested here. In an ongoing study in our lab we are looking at older adults with presbycusis to determine whether the pattern of correlations reported here is modified by age, hearing status and hearing aid use, and whether rapid learning is associated with clinically relevant indices of speech perception (preliminary outcomes have been submitted by Rotman *et al.* (2020) to these proceedings). If our hypothesis is correct, it follows that rapid learning might be one of the factors partially predicting how well listeners will adapt to new hearing aids. We acknowledge that none of these proposals can provide definitive proof of our hypothesis, but negative findings can certainly falsify it.

Our hypothesis was driven to a great extent by our frustration with the literature on the potential clinical application of auditory training (Pichora-Fuller and Levitt, 2012). Specifically, we maintain that neither bottom-up nor top-down approaches were successful in broadening the scope of learning generalization. Nevertheless, although auditory training studies failed to prove effective under rigorous scrutiny (Henshaw and Ferguson, 2013; Saunders *et al.,* 2016), there are individuals who report training-related benefits (see Lavie *et al.,* 2013; Karawani *et al.,* 2016). It is therefore interesting to end with speculating that perhaps, for some individuals, training may have pushed rapid perceptual learning to support speech perception under ecological conditions, rather than have a direct impact on speech perception. Testing this speculation requires further studies.

ACKNOWLEDGEMENTS

This study was supported by the Israel Science Foundation grant 206/18. We thank our 2017-2018 audiology students who collected the data for this manuscript as part of their undergraduate research projects.

REFERENCES

Adank, P. and Janse, E. (**2009**). “Perceptual learning of time-compressed and natural fast speech,” J. Acoust. Soc. Am., **126**, 2649-2659. doi: 10.1121/1.3216914

Altmann, T. and Young, D. (**1993**). “Factors affecting adaptation to time-compressed speech,” EUROSPEECH '93. Berlin**,** 333-336.

Banai, K. and Lavner, Y. (**2014**). “The effects of training length on the perceptual learning of time-compressed speech and its generalization,” J. Acoust. Soc. Am., **136**, 1908. doi: 10.1121/1.4895684

Burk, M. H. and Humes, L. E. (**2008**). “Effects of long-term training on aided speech-recognition performance in noise in older adults,” J. Speech Lang. Hear. Res., **51**, 759-771. doi: 10.1044/1092-4388(2008/054)

Dupoux, E. and Green, K. (**1997**). “Perceptual adjustment to highly compressed speech: Effects of talker and rate changes,” J. Exp. Psychol. Human, **23**, 914-927. doi: 10.1037/0096-1523.23.3.914

Green, C. S., Banai, K., Lu, Z. and Bavelier, D. (**2019**). “Perceptual Learning,” *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. J. T. Wixted, John Wiley & Sons, Inc.**,** 1-47. doi: 10.1002/9781119170174.epcn217

Henshaw, H. and Ferguson, M. A. (**2013**). “Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence,” PLoS One **8**, e62836. doi: 10.1371/journal.pone.0062836

Karawani, H., Bitan, T., Attias, J. and Banai, K. (**2016**). “Auditory perceptual learning in adults with and without age-related hearing loss,” Front Psychol., **6**, 2066. doi: 10.3389/fpsyg.2015.02066

Karawani, H., Lavie, L. and Banai, K. (**2017**). “Short-term auditory learning in older and younger adults,” Proc. ISAAR, **6**, 1-8.

Lavie, L., Attias, J. and Karni, A. (**2013**). “Semi-structured listening experience (listening training) in hearing aid fitting: influence on dichotic listening,” Am. J. Audiol., **22**, 347-350. doi: 10.1044/1059-0889

Manheim, M., Lavie, L. and Banai, K. (**2018**). “Age, Hearing, and the Perceptual Learning of Rapid Speech,” Trends Hear., **22**, 2331216518778651. doi: 10.1177/2331216518778651

Mattys, S. L., Davis, M. H., Bradlow, A. R. and Scott, S. K. (**2012**). “Speech recognition in adverse conditions: A review,” Lang. Cogn. Process., **27**, 953-978. doi: 10.1080/01690965.2012.705006

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**, 1315-1330. doi: 10.1037/0096-1523.31.6.1315

Pichora-Fuller, M. K. and Levitt, H. (**2012**). “Speech comprehension training and auditory and cognitive processing in older adults,” Am. J. Audiol., **21**, 351-357. doi: 10.1044/1059-0889

Prior, A. and Bentin, S. (**2006**). "Differential integration efforts of mandatory and optional sentence constituents," Psychophysiology, **43**, 440-449. doi: 10.1111/j.1469-8986.2006.00426.x

Ronnberg, J., E. Holmer and Rudner, M. (**2019**). "Cognitive hearing science and ease of language understanding," Int. J. Audiol., **58**, 247-261. doi: 10.1080/14992027.2018.1551631

Rotman, T., Lavie, L., and Banai, K. (**2020**). "Rapid perceptual learning of time-compressed speech and the perception of natural fast speech in older adults with presbycusis," Proc. ISAAR, **7**, 93-100.

Samuel, A. G. and Kraljic, T. (**2009**). "Perceptual learning for speech," Atten. Percept. Psychophys., **71**, 1207-1218. doi: 10.3758/APP.71.6.1207

Saunders, G. H., Smith, S. L. , Chisolm, T. H. , Frederick, M. T. , McArdle, R. A. and Wilson, R. H. (**2016**). "A randomized control trial: Supplementing hearing aid use with listening and communication enhancement (LACE) auditory training," Ear Hearing, **37**, 381-396. doi: 0.1097/AUD.0000000000000283

Sweetow, R. W. and Sabes, J. H. (**2006**). "The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program," J. Am. Acad. Audiol., **17**, 538-558. doi: 10.3766/jaaa.17.8.2

Verhelst, W. and Roelands, M. (**1993**). "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Minneapolis, MN, USA**,** 554-557.

Wright, B. A., Wilson, R. M. and Sabin, A. T. (**2010**). "Generalization lags behind learning on an auditory perceptual task," J. Neurosci. **30**, 11635-11639. doi: 10.1523/JNEUROSCI.1441-10.2010