More than adaptation – evidence for training-induced perceptual learning of time-compressed speech

KAREN BANAI[1,*] AND YIZHAR LAVNER[2]

[1] *Department of Communication Sciences and Disorders, University of Haifa, Haifa 31905, Israel*

[2] *Department of Computer Science, Tel Hai College, Tel Hai 12208, Israel*

The identification of time-compressed speech improves significantly following short-term exposure, but it is not clear whether additional practice yields additional learning. The goal of the experiment reported here was to determine whether 30-40 minutes of training, during which listeners practiced the identification of 100 different time-compressed sentences, yielded additional learning to that induced by a single brief exposure to 20 sentences. We also asked if this learning generalized to novel sentences and to a new speaker. Training resulted in more learning than a single brief exposure, and this learning generalized to a new speaker but not to new tokens. Brief exposure to 20 sentences did not result in any significant increases to performance when compared to naive listeners. We conclude that a prolonged learning phase exists for time-compressed speech, but that learning during this phase does not fully transfer to new, untrained tokens.

INTRODUCTION

The identification of time-compressed speech, an artificially created form of rapid speech, improves rapidly with exposure to a few time-compressed sentences, a phenomenon referred to as adaptation (e.g., Sebastian-Galles *et al.*, 2000; Peelle and Wingfield, 2005) or perceptual adjustment (e.g., Dupoux and Green, 1997; Pallier *et al.*, 1998). However, whether learning beyond this brief adaptation phase also occurs, and if so whether its characteristics are distinct from those of initial adaptation, remains unclear, because systematic training on more than 10-20 stimuli has been rare. Consistent with the finding that even highly experienced non-native speakers benefit from slower than normal presentation (Conrad, 1989; Zhao, 1997), we have previously observed a prolonged learning phase on a time-compressed speech identification task among non-native speakers of Hebrew (Banai and Lavner, 2012). The goal of the experiment presented here was to extend these findings to the learning of time-compressed speech in native speakers.

Relatively brief adaptation (10-20 sentences) to time-compressed speech substantially improves its perception, albeit not perfectly so. Previous reports suggest that after such exposure, performance improves from 20-76% correct to the range of 40-85% correct for the level of compression used during adaptation (Altmann and Young, 1993; Dupoux and Green, 1997; Pallier *et al.*, 1998;

*Corresponding author: kbanai@research.haifa.ac.il

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

Sebastian-Galles *et al.*, 2000; Golomb *et al.*, 2007; Adank and Janse, 2009). In these studies, improvement appears quite general in the sense that transfer was observed across stimuli and even across languages. However, it is not clear whether performance does not reach ceiling (100% correct) due to inherent limitations of the learning process or due to other factors such as the duration of training. In a previous study we have shown that the latter might be one of the reasons (Banai and Lavner, 2012). In this study, we trained non-native speakers of Hebrew on the semantic verification of time-compressed sentences using an adaptive procedure. Listeners improved significantly over the course of a training program in which they had to verify 300 sentences per session for five sessions. Post training, the ability of trained listeners, but not of untrained controls who participated in pre- and post-test sessions only, to verify the trained sentences, became as good as that of naive native speakers. Both trained non-native listeners and naive native ones were able to consistently verify sentences compressed to less than 30% of their original length. The effects of learning generalized to the identification of time-compressed sentences produced by different talkers but not to untrained sentences or single words, leading us to hypothesize that prolonged learning might constitute a different, more specific form of learning than that observed after brief adaptation.

Although our previous study suggests that prolonged learning does occur on time-compressed speech, we could not document its presence among native speakers. During the pre-test phase native listeners could consistently verify sentences compressed to the maximum level of compression we allowed the adaptive procedure to reach during this phase (20%), making it impossible to uncover any further learning. It is thus still possible that the prolonged learning observed among non-native speakers simply reflect slower adaptation. Therefore, we now ask whether the identification of time-compressed speech continues to improve beyond the effects of adaptation to 20 sentences in native speakers of Hebrew. We also ask whether the pattern of generalization is similar or different from that observed among non-native speakers and after brief adaptation.

METHODS

Participants

Thirty native Hebrew speakers participated in this experiment. All were undergraduate University of Haifa students, with no history of hearing, learning, or language problems. Participants were paid for their participation. All aspects of the study were approved by the ethics committee of the Faculty of Social Welfare and Health Sciences at the University of Haifa. Participants were divided into three groups: a trained group (n = 10), an untrained control group (n=10), and a group of naive listeners (n = 10).

Organization of the experiment

The experiment had three phases. A pre-test on which trained listeners and untrained controls were exposed to 20 sentences (taken from the training set) presented by a male speaker and compressed to 30% of their naturally spoken duration; a training session during which trained listeners practiced a compressed speech verification task during which the degree of compression varied adaptively based on performance; and a post-test on which trained listeners, untrained controls (who attended the pre-test but not the training session) and naive listeners (who participated only in the post-test) were exposed to the 20 sentences from the pre-test as well as the same 20 sentences presented by a different speaker and 20 novel sentences spoken by the trained speaker, all compressed to 30%.

Stimuli

Stimuli were simple active subject-verb-object sentences in Hebrew, each 5-6 words long, taken from Prior and Bentin (2006). A total of 120 sentences were used. One hundred sentences formed the training set. The other 20 were used to assess the generalization of learning to untrained tokens. Sentences were recorded and then compressed with a WSOLA algorithm (Verhelst and Roelands, 1993) implemented in Matlab. For further details see Banai and Lavner (2012).

Tasks

Pre- and post-tests. Sentences were presented in blocks of 20 sentences. After hearing each sentence, listeners were asked to write it down as accurately as they could. No feedback was provided during this phase.

Training. Five blocks of 60 sentences selected at random from the training set were presented. On each trial listeners had to determine whether the sentence they heard was semantically plausible (e.g., 'the grumpy waiter served the soup') or implausible (e.g., 'the grumpy potato served the soup'). Initial compression level was 65%. Subsequently, the level of compression changed based on listeners response using a 2-down/1-up staircase procedure. Feedback was provided for both correct and incorrect responses. For further details see Banai and Lavner (2012).

Experimental Conditions

The trained condition was comprised of 100 sentences presented by a male talker (designated the trained talker).

Three additional conditions were used:

1) Trained tokens: 20 sentences, randomly selected from the training set presented by the trained talker. These were presented to trained and control listeners during the pre- and post-test phases and to naive listeners during the post-test.
2) Untrained tokens: 20 sentences, not included in the training set, presented by the trained talker. These were administered during the post-test only to all listeners.
3) Untrained talker: The 20 trained tokens from above presented by a different speaker. These were administered during the post-test only to all listeners.

RESULTS

Training-induced learning

To determine whether additional learning to that induced by exposure to 20 sentences occurred, the learning curves from the 5 blocks of the training phase were analyzed. The mean group and the individual learning curves are shown in Fig. 1. The slopes of 9/10 individual curves were negative, indicating that in general, thresholds tended to improve with practice. The mean slope (−0.013 ± 0.01) was significantly negative with a 95% confidence interval of −0.019 to −0.006. A repeated measures ANOVA over the training blocks with contrasts comparing the mean of each block to the mean of all previous blocks suggests that learning was evident starting the third block and continued through the fourth block of training ($F(4,36) = 7.69$, $p < 0.001$; block 2 vs. block 1: $F(1,9) = 2.14$, $p = 0.18$; block 2 vs. previous blocks: $F(1,9) = 6.45$, $p = 0.032$; block 4 vs. previous blocks: $F(1,9) = 25.56$, $p = 0.001$; block 4 vs. previous: $F(1,9) = 4.49$, $p = 0.063$). Together it therefore appears that significant perceptual learning on a time-compressed speech task continues even beyond the initial adaptation phase.

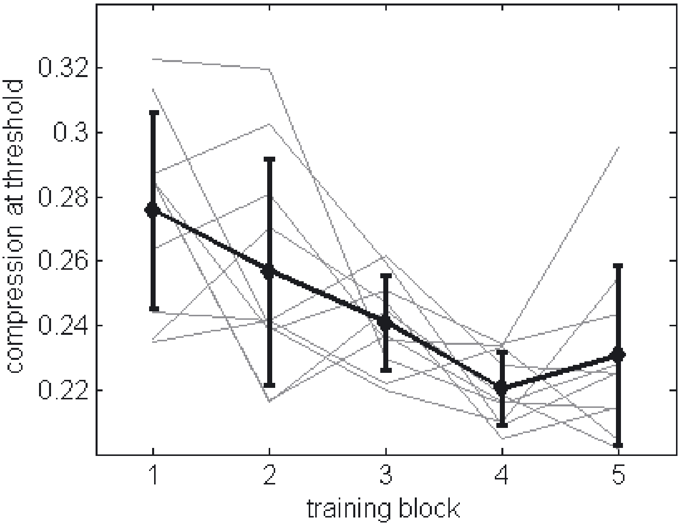

Fig. 1: Mean group learning curve (thick line) and individual curves (thin lines). Thresholds were determined as the mean compression over the last 5 reversals in a given block of trials.

Training versus rapid-learning

To determine whether the training-induced learning (discussed above) was significantly greater than the rapid learning induced by participating in the pre- and post-test only, performance on the trained tokens was compared between trained and control listeners. As shown in Fig. 2, pre- to post-test improvement was greater in the trained than in the control group. A significant group × test session interaction in an ANOVA with session as within subject factor and group as a between factor one ($F(1,17) = 10.69$, $p = 0.005$, partial $\eta^2 = 0.39$) suggested that the improvement was significantly greater among trained listeners. Likewise, an analysis of co-variance (ANCOVA, with pre-test scores as covariate) on the percentage of words correctly recognized during the post-test also suggests that mean post-training performance is significantly better among trained listeners even after taking into account putative pre-test differences ($F(1,18) = 31.84$, $p < 0.001$, partial $\eta^2 = 0.66$). Significant generalization of learning to the untrained talker condition ($F(1,18) = 8.47$, $p = 0.010$, partial $\eta^2 = 0.35$), but not to the untrained tokens condition ($F(1,18) = 0.38$, $p = 0.54$, partial $\eta^2 = 0.02$) was also observed.

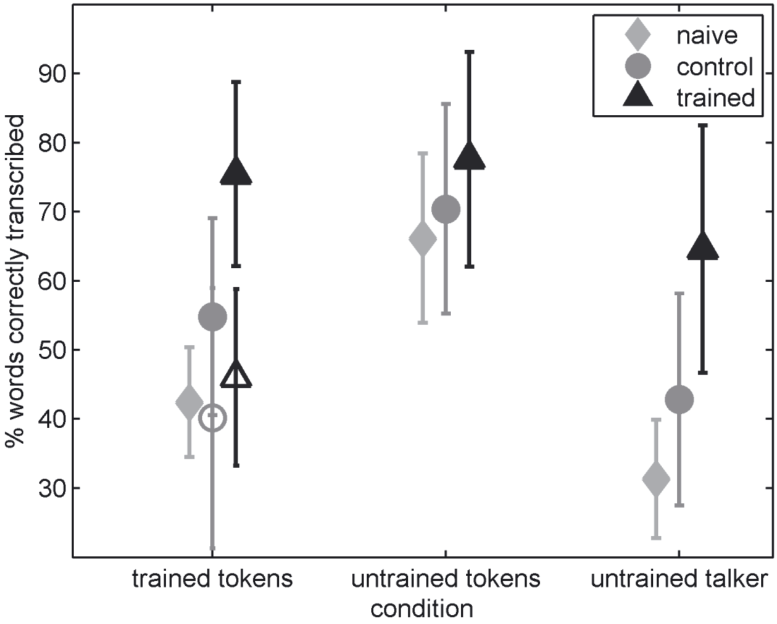

Fig. 2: Mean group performance across conditions and test sessions. Empty symbols denote pre-test performance; filled symbols denote post-test and naive performance. Error bars are ±1 standard deviation.

Training and pre-test participation versus naive performance

Finally, post-test performance was compared across the 3 group of listeners. ANOVAs suggest group differences on two of the three conditions (trained tokens: $F(2,26) = 31.46$, $p < 0.001$; untrained talker: $F(2,26) = 13.91$, $p < 0.001$), but not on the untrained tokens condition. To determine whether significant group differences were associated with the initial learning that occurred during the pre-test or with the effects of training, planned comparisons were carried out. Controls and naive listeners performed similarly on all conditions (see Fig. 2), suggesting that pre-test participation did not yield meaningful gains relative to naive performance. On the other hand, controls were significantly poorer than trained listeners on both the trained tokens condition ($t = 6.14$, $p < 0.001$) and the untrained talker condition ($t = 3.18$, $p = 0.004$). Together these suggest that group differences arose because training yielded more benefits than either pre-test or post-test participation, and pre-test induced learning had no lasting effects when compared with naive performance.

DISCUSSION

Similar to our previous findings in non-native speakers (Banai and Lavner, 2012), we now report that prolonged, training-induced learning of time-compressed speech occurs also among native speakers of Hebrew. Learning generalized to a novel talker of the opposite sex, but appeared specific to the sentences encountered during training. This pattern of learning and generalization suggests that training-induced learning might be more specific than adaptation-induced gains. Therefore, we conclude that adaptation-induced and training-induced learning of time-compressed speech might represent different types or phases of learning, such that learning starts with a rapid and broad phase during which learning generalizes quite widely and continues with a slower and more stimulus specific phase. This conclusion is consistent with the Reverse Hierarchy Theory (RHT) of perceptual learning (e.g., Ahissar *et al.*, 2009).

In contrast to the present findings, studies on the rapid adaptation to time-compressed speech suggest that learning during this phase is not stimulus-specific. For example, the ability of listeners to reproduce a specific time-compressed sentence was better if this sentence was encountered after 10 other sentences than if that same sentence was encountered after 5 other sentences (Dupoux and Green, 1997). Likewise, listeners who adapted to a set of 10 sentences in Catalan, reported a set of test sentences in Spanish more accurately than un-adapted controls (Pallier *et al.*, 1998). This was true even for listeners who spoke no Catalan, providing another demonstration of the generality of learning in this rapid phase. Generalization after training-induced learning in the current study was more limited. The lack of generalization to novel sentences in particular suggests that the learning we observed was different in nature than that reported in earlier studies. Otherwise trained listeners, who recognized a subset of the trained sentences more accurately than naive ones, would have also been more accurate on the set of untrained tokens.

ACKNOWLEDGEMENTS

This study was supported by the National Institute of Psychobiology in Israel.

We thank Yael Braver, Qamar Daher, Alex Frid, Dorit Nesher, and Neta Yuval-Weiss for help with various aspects of this project.

REFERENCES

Adank, P. and Janse, E. (**2009**). "Perceptual learning of time-compressed and natural fast speech," J. Acoust. Soc. Am., **126**, 2649-2659.

Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (**2009**). "Reverse hierarchies and sensory learning," Philos. Trans. R. Soc. Lond. B. Biol. Sci., **364**, 285-299.

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

Banai, K., and Lavner, Y. (**2012**). "Perceptual learning of time-compressed speech: more than rapid adaptation," PLoS One, **7**, e47099.

Conrad, L. (**1989**). "The effects of time-compressed speech on native and EFL listening comprehension," Stud. Second Lang. Acq., **11**, 1-16.

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.

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.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). "Perceptual adjustment to time-compressed speech: a cross-linguistic study," Mem. Cognit., **26**, 844-851.

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. Human, **31**, 1315-1330.

Prior, A., and Bentin, S. (**2006**). "Differential integration efforts of mandatory and optional sentence constituents," Psychophysiology, **43**, 440-449.

Sebastian-Galles, N., Dupoux, E., Costa, A., and Mehler, J. (**2000**). "Adaptation to time-compressed speech: Phonological determinants," Percept. Psychophys., **62**, 834-842.

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.

Zhao, Y. (**1997**). "The effects of listeners' control of speech rate on second language comprehension," Appl. Linguist., **18**, 49-68.