Tasks, Talkers and the Perceptual Learning of Time-Compressed Speech

Maram Tarabeih- Ghanayim

University of Haifa

Yizhar Lavner

Tel-Hai College

Karen Banai

*University of Haifa**Author Note*

Maram Tarabeih-Ghanayim and Karen Banai, Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel; Yizhar Lavner, Department of Computer Science, Tel-Hai College, Tel-Hai, Israel

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

Correspondence concerning this article should be addressed to Karen Banai, Department of Communication Sciences and Disorders, University of Haifa, Haifa, 3498838, Israel. Email: kbanai@research.haifa.ac.il

NOTE: This is the preprint of: Tarabeih-Ghanayim, M., Lavner, Y., & Banai, K. (2020). [Tasks, Talkers, and the Perceptual Learning of Time-Compressed Speech](#). *Auditory Perception & Cognition*, DOI: [10.1080/25742442.2020.1846011](https://doi.org/10.1080/25742442.2020.1846011)

The final version has been substantially revised from this preprint!

Abstract

Many auditory skills (e.g., the recognition of perceptually difficult speech) improve with practice, but the generalization of this learning to untrained materials is limited. Here, we asked whether the type of practice (semantic or accent judgment) and talker variability (defined as the number of different talkers encountered during practice, two or six), influenced the perceptual learning of time-compressed speech and its generalization to unpracticed materials. Four groups of participants trained on the four task/talker variability combinations, and their pre- and post-training recognition of time-compressed speech was compared to that of a group of untrained participants ($n = 14-16$ participants/group). Across groups, training led to substantial learning of the trained tokens and to generalization to new talkers producing previously encountered sentences (compared to the untrained control group). However, neither type of training had a significant effect on the recognition of new sentences, even for familiar talkers. Semantic training yielded more learning and better retention of learning over a 2 week interval than accent training. The number of talkers had only marginal effects. These results suggest that learning of time-compressed speech is robust and only partially task specific, but its generalization to untrained tokens following brief practice is limited. In contrast to other types of speech training, here exposure to a variety of talkers during training did not contribute to the transfer of learning to new materials.

Tasks, Talkers and the Perceptual Learning of Time-Compressed Speech

The perception of speech, especially under acoustically difficult conditions is subject to perceptual learning (Green, Banai, Lu, & Bevelier, 2018; Samuel & Kraljic, 2009). However, this learning is often specific to the stimuli encountered during practice, with little generalization to untrained materials (Banai & Lavner 2014, 2019; Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011; Manheim, Lavie, & Banai, 2018). The factors limiting the transfer of perceptual learning, especially under more naturalistic circumstances (e.g., as in the case of perceptually difficult speech, such as rapid speech and speech in noise) are not well understood. Understanding the factors that may facilitate or hinder generalization is important theoretically and for potential clinical and educational application of training programs that are based on perceptual learning. To mimic some of the challenges created by rapid speech rates, we used time-compressed speech for both training and testing. We focused on two factors, talker variability and training task, which were shown to contribute to the transfer of perceptual learning to new contexts (Lively, Logan, & Pisoni, 1993; Loebach, Bent, & Pisoni, 2008). We assessed transfer to new sentences (semantic transfer), to new talkers presenting previously encountered sentences (acoustic transfer) and to new sentences presented by a novel talker (acoustic-semantic transfer).

The Specificity of the Perceptual Learning of Acoustically Modified Speech

Studies on the perceptual learning of speech made hard to perceive due to spectral (e.g., noise vocoding) or temporal (e.g., time-compression) manipulations, suggest that while the perception of these forms of speech readily improves with practice, this learning is at least partially specific to the acoustic characteristics encountered during training (Banai & Lavner 2014; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Hervais-Adelman et al., 2011; Loebach et al.,

2008). For example, learning filtered noise-vocoded speech generalizes completely between different frequency regions, which means listeners trained to perceive high-pass filtered vocoded speech are also able to perceive new low-pass filtered vocoded speech (and vice-versa for listeners trained with low-pass vocoded speech) (Hervais-Adelman et al., 2011). On the other hand, the transfer of perceptual learning of vocoded speech from one carrier signal (noise bands, pulse trains, and sine waves) is not complete, suggesting that the acoustic characteristics of the signal may impact the transfer of learning (Dahan & Mead, 2010; Hervais-Adelman et al., 2011). Dahan & Mead (2010) found that training led to generalization to new noise-vocoded words presented by the same speaker, and this generalization increases as the phonological similarity between the trained and novel utterances increases. Nevertheless, generalization to a new speaker was inconsistent. Surprisingly, greater cross-speaker generalization was observed on stimuli where speakers were acoustically more different, such as male and female (Dahan & Mead, 2010).

Although abrupt acoustic changes (in talker characteristic or speech rate) had little effect on immediate adjustment to time-compressed speech (Dupoux & Green, 1997; Simhony, Grinberg, Lavie, & Banai, 2014), we hypothesize that the acoustic characteristics of the signal (dictated by the anatomy and physiology of the vocal apparatus of individual talkers) do play a role in the transfer of learning. This hypothesis is driven by two reasons: First, the pattern of transfer between time-compressed and natural-fast speech is asymmetrical, with acoustic-semantic transfer from time-compressed to natural-fast speech, but not the other way around (Adank & Janse, 2009). Second, we have recently reported that although rapid learning of time-compressed speech (tested similarly to the adjustment protocols discussed above) resulted in some acoustic-semantic transfer, additional training-induced gains were more specific, with

no acoustic-semantic transfer to new natural-fast speech (Manheim et al., 2018). Whether longer training, which was reported to support semantic transfer (Banai & Lavner 2014), also supports acoustic-semantic transfer is unknown.

Another factor investigated in the context of the specificity of perceptual learning is task specificity (Hawkey, Amitay, & Moore, 2004; Polley, Steinberg, & Merzenich, 2006). By one account (the reverse hierarchy theory, RHT), attention to the specific acoustic dimension dictated by the task performed during training, results in learning that is specific to the trained task with little transfer to different dimensions (Ahissar, Nahum, Nelken, & Hochstein, 2009). Consistent with the RHT, when rats were trained to attend to a task- relevant acoustic feature (either frequency or intensity), both learning and changes in the cortical representation of the trained features were task specific even though that acoustic characteristic of the sounds to which the rats were exposed to during training were similar across groups (Polley et al., 2006). Similar behavioral findings were reported in humans (Halliday, Moore, Taylor, & Amitay, 2011; Loebach et al., 2008). In one study, vocoded speech was presented with different training tasks (gender identification, talker identification, speech recognition), and the transfer of learning was tested (Loebach et al., 2008). Subjects trained on talker identification performed similarly to those trained on speech recognition (with new sentences and sentences that were more spectrally degraded), and significantly better than the subjects from the gender identification group. The writers suggested that training on explicit indexical (relating to the characteristics of the speaker's voice such as gender, age and dialect) tasks can yield as much perceptual learning and transfer as training on traditional recognition tasks if task demands are high enough to require sustained attention and deeper processing (the case of talker but not gender identification). With time-compressed speech, previous findings suggest limited

transfer even when tasks differed in high-level task demands and not in the attended acoustic dimension (Banai & Lavner, 2014, 2019). In these studies, training was conducted using a semantic verification task in which listeners had to indicate whether a time-compressed sentence was semantically plausible or not, but tested on a speech recognition (sentence transcription) task.

Talker Variability and Perceptual Learning of Speech

Although acoustic specificity might be a hallmark of perceptual learning (see Green et al., 2018), one factor that potentially mitigates the degree of specificity in speech learning is talker variability (i.e., increasing the number of different talkers experienced during training, Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Lively et al., 1993). Increasing the number of talkers modifies the acoustic characteristics of the trained materials, without changing their lexical or semantic content (Goldinger, Pisoni, & Logan, 1991). In a seminal study, the contribution of talker variability to the categorical learning of a foreign phonemic contrast was studied. Japanese listeners were taught to discriminate the /r/-/l/ contrast which is not a phonemic contrast in Japanese. Participants were trained on a minimal pair identification task that was presented either by five different talkers or by one single talker. Training with five different talkers led to more transfer to new words produced by a familiar/ unfamiliar talker than practice with a single talker (Lively et al., 1993). This finding has been replicated several times since and shows that the high-variability training encourages a long-term modification of listeners' phonetic perception (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994).

Another area in which talker variability was found to promote the transfer of learning across talkers is the perception of accented speech. Thus, although mixing a

number of talkers within a stimulus set was reported to impair speech recognition (at least under difficult conditions; Bent & Frush Holt, 2013; Martin, Mullennix, Pisoni, & Summers, 1989; McLennan & Luce, 2005), a few studies reported that talker variability supports the learning of new accented speech (Brosseau-Lapr, Rvachew, Clayards, & Dickson, 2013; Sidaras, Alexander, & Nygaard, 2009). One study that investigated the effect of talker variability on the categorization of regional dialects found that talker variability disturbed the learning process but enhanced the transfer to a new talker (Clopper & Pisoni, 2004). Another study (Bradlow & Bent, 2008) found that talker variability contributed to the transfer of learning across talkers that shared the same accent of the talkers encountered in training, but not to a new (different) accent. Later, Baese-Berk et al., (2013) found that training with multiple- talkers from variable accents led to transfer to a new talker from a new accent. Cross-talkers transfer probably depends on the similarity between the trained and the new talkers (Kraljic & Samuel, 2005; Reinisch & Holt, 2014). On the other hand, a study which investigated the recognition and learning of monosyllabic English words produced by 4 different native Spanish speakers found that high-variability training did not contribute to the recognition of new words produced neither by known nor by novel speaker (Wade, Jongman, & Sereno, 2007). Recent studies on perceptual learning of time-compressed speech with a single talker have shown robust learning and partial acoustic or semantic transfer (Banai & Lavner 2014, 2016, 2019), but there was no acoustic-semantic transfer reported (Manheim et al., 2018). Therefore, one of the goals of the current study was to determine if talker variability can promote the transfer of learning of time-compressed speech to novel tokens presented by new talkers.

The Current Study

This study aims to investigate the contribution of the two factors discussed above, training task and talker variability, to the perceptual learning of time-compressed speech and its acoustic and semantic transfer. To this end, during training, different groups of listeners were trained on two different tasks (a semantic task requiring semantic verification or an ‘acoustic’ task requiring accent verification) with either high or low talker variability. Prior to training, baseline recognition of time-compressed speech was assessed. After training (immediately and after two weeks), participants were tested on a series of conditions intended to estimate the transfer of learning. During baseline and test, participants had to transcribe what they heard as accurately as they could.

Based on previous studies on time-compressed speech, we hypothesized that *all types of training will result in both learning and transfer*, compared to untrained participants. The four training groups were exposed to time-compressed sentences several times during training because according to the RHT, stimulus repetitions may direct attention to task-relevant acoustic parameters and thus enable the use of low-level information relevant to the processing of time-compressed speech (Ahissar et al., 2009). However, while learning is expected across conditions, *training on the semantic task is expected to yield more learning and transfer than training on the acoustic task*. According to the transfer appropriate processing (TAP) theory of learning and memory, performance will be maximized when the task used during testing is similar to the task used during training (Morris, Bransford, & Franks, 1977). In our study, the transcription task used in the tests is more similar to the semantic task than to the acoustic task, because both of them (the transcription task and the semantic task) rely on the extraction of lexical information. Still, we decided to train two groups on an accent verification task to determine the influence of attention to this more acoustic dimension during

training on perceptual learning of time-compressed speech. Note that training on an indexical task that require high-level attention and deep processing can lead to high levels of perceptual learning and transfer (Loebach et al., 2008).

Lastly, based on the literature reviewed above, we hypothesize that *across tasks, talker variability will interrupt learning, resulting in less learning on the trained conditions. Additionally, talker variability is expected to enhance transfer following semantic training.* According to the RHT, training promotes the ability to attend to task-relevant acoustic cues. Talker variability may slow or interfere with this process because it reduces the consistency of acoustic characteristics during training. In this case, the ability of listeners to utilize the appropriate low-level information will be impaired and thus learning will be reduced regardless of the training task (Nahum, Nelken & Ahissar, 2010). On the other hand, when required to learn higher-level information which is not talker-specific, as in the semantic task, talker variability may prove conducive of transfer, because learning on the semantic task required focus on high-level cues even in the presence of the acoustic challenge created due to time compression.

Methods

Participants

Eighty native Hebrew speakers (ages 18-35 years, $M = 25$, $SD = 3$) participated in this study. All participants reported having no history of hearing, learning, language or attention problems and had no prior experience with time-compressed speech. Participants were paid for their participation (\$15-20). 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 (permit # 199/12).

Participants were divided randomly to five groups, an untrained control group and four groups that received training on one of the four training protocols described below (see Table 1).

Table 1. Group characteristics

Group	Number of participants ¹	Males	Females	Days between sessions ²
1) Untrained control	16	2	14	16±3
2) low-variability, acoustic	15	2	13	15±2
3) high-variability, acoustic	16	3	13	16±2
4) low-variability, semantic	16	1	15	15±2
5) high-variability, semantic	17	4	13	15±2

¹ The number of participants in the baseline and in the immediate test.

² Average and standard deviation (rounded to nearest full day).

Stimuli

140 simple Hebrew sentences (taken from Prior and Bentin, 2006) were recorded. Sentences were five to six words long and had a common subject-verb-object grammatical structure which is prevalent in Hebrew. Half the sentences were semantically plausible (for example: "The famous soloist played a classical guitar"), while the other half were semantically implausible ("The sensitive painter fell in love with a thick rope"). Before recording, the talkers reviewed the sentences, which was particularly necessary for the non-native talkers. Consequently, they were asked to read a sequence of sentences clearly and at a natural rate. The sentences were recorded at a sampling rate of 44 kHz, using a standard microphone, in a sound attenuating booth

and were edited with Audacity audio editor. The editing included removal of background noise (computer noise, movements of the talkers) and cutting each sentence and then saving it as a separate sound file. The Sentences were time-compressed with a WSOLA time-scale modification algorithm (Verhelst & Roelands, 1993) and were presented to participants via a PC computer and circumaural headphones (Sennheizer HD-215). Sentences were recorded by five native Hebrew talkers and five non-native talkers with different accents: Arab, Russian and American. The high variability training set included 100 sentences recorded by six speakers (half of them are non-native), and the low variability training set was performed by two speakers only (one is native and the other is non-native). The 40 remaining sentences were used to transfer and retention tests: 20 sentences were recorded by one of the native talkers from the training (half of them were used in the transfer tests and the other half were used in the retention test). The remaining 20 sentences were recorded by two new native talkers (one of them was used for transfer tests and the other one for the retention test). The characteristics of each talker are detailed in Table 2.

A pilot experiment was conducted to assess the intelligibility of the talkers by Hebrew-speakers with normal hearing. Talkers with High, medium and low levels of intelligibility and clarity were selected.

Table 2. Speakers' characteristics

Speaker	Sex	First language	Speech rate ¹	F0 ²	Formants ³		
					F1	F2	F3
1	Male	Hebrew	106±11	135±15	587±35	1325±29	2469±81

Speaker	Sex	First language	Speech rate ¹	F0 ²	Formants ³		
					F1	F2	F3
2	Female	Hebrew	102±14	182±34	733±54	1559±278	2752±471
				116-80			
3	Male	Hebrew	123±12	107±18	542±21	1296±53	2554±79
				84-176			
4	Male	Arabic	115±15	126±14	696±19	1314±40	2185±53
				98-178			
5	Male	Russian	136±12	124±8	487±30	1316±48	2561±115
				105-50			
6	Female	English	92±14	182±32	714±53	1370±112	2427±263
				102-53			
7	Male	Hebrew	120±12	119±22	700±21	1332±37	2391±55
				80-173			
8	Female	Arabic	100±14	208±32	820±42	1451±162	2097±238
				99-276			
9	Male	Hebrew	105±12	104±13	587±30	1258±40	2532±141
				84-143			
10	Female	Arabic	93±11	213±28	925±58	1548±82	2111±130
				105-83			

¹ Each cell shows mean ± SD in words/minute.

² Fundamental frequency (Hz), each cell shows mean ± SD and range.

³ Formant values in Hz, mean ± SD are shown in each cell.

Overall Design

The experiment comprised four phases: A baseline, a training phase (for trained participants only), an immediate test and a delayed test that was performed two to three weeks after the immediate test. The first three phases were performed on the same day. The control group had received a half-hour break between the baseline and the immediate test, a duration similar to that of the training phase (see Table 3).

Table 3. Experimental procedure and conditions

Group	Baseline ¹	Training ²	Testing (Immediate and delayed) ³
1) Untrained control	√	X	√
2) low-variability, acoustic	√	accent, 2-talkers	√
3) high-variability, acoustic	√	accent, 6-talkers	√
4) low-variability, semantic	√	semantic, 2-talkers	√
5) high-variability, semantic	√	semantic, 6-talkers	√

¹ 10 sentences from the training set.

² five blocks of 60 sentences each.

³ five conditions of 10 sentences per test: (1) repetition of the baseline; (2) new sentences; (3) new native talker; (4) new non-native talker; (5) new sentences and new talker.

Baseline and Test Conditions

All participants performed a baseline (pre-test) and an immediate test (comprised of five test conditions, see Table 3) on the same day. Test conditions included one test which is similar to the baseline, intended to assess learning of the trained stimuli and four conditions designed to assess the transfer of learning. Each test condition included 10 sentences compressed to 30% of their naturally spoken duration. Half of the sentences in each condition were semantically plausible while the others

were semantically implausible. Upon hearing a sentence, listeners were asked to write it down as accurately as they could. The details of each test are described below. The learning condition was administered first, followed by the transfer conditions which were presented at a random but uniform order across participants.

Baseline. This test included ten sentences taken from training sentences. Sentences were recorded by a native speaker of Hebrew (speaker 1) who was rated "medium" in terms of intelligibility and clarity.

Immediate test. Included tests of learning (defined as performance with tokens repeated from previous phases) and transfer (defined as performance with tokens that were different from previously encountered stimuli acoustically, semantically or both acoustically and semantically).

Learning. Same ten sentences and same talker as in the baseline test.

Semantic transfer. Ten new sentences that were not presented during training or other transfer tests, all recorded by speaker 1.

Acoustic transfer - native talker. Ten sentences taken from training sentences (and were not presented during other transfer tests). Sentences were recorded by a new native Hebrew speaker (speaker 7), rated "medium" in terms of intelligibility and clarity.

Acoustic transfer - non-native talker. Ten sentences taken from training sentences (not presented during other transfer tests). Sentences were recorded by a new non-native Hebrew speaker of Hebrew (speaker 8) who had "medium" intelligibility and clarity.

Acoustic-semantic transfer. 10 new sentences that were not presented during training or other transfer tests. Sentences were recorded by a native Hebrew (speaker

7) that did not record the training sentences and was rated as "medium" in terms of intelligibility and clarity.

Delayed test. Seventy-six participants (out of 80 who attended the immediate tests) participated in the delayed test. This test was administered two to three weeks ($M = 15$, $SD = 2$ days) after the first session which included the baseline, training and immediate test. This phase included a repetition of the baseline test. Additionally, participants performed the same transfer tests as in the immediate test, but with different sentences and (in the acoustic transfer conditions) new speakers (speakers 9+10).

Training Phase and Procedure

Training was administered in five blocks of 60 sentences completed over the course of approximately 40 minutes (The sentences were selected randomly from a total of 100 training sentences). Training started at a compression level of 0.65. Subsequently, compression was adapted using a 2-down/1-up staircase procedure with a maximal compression of 0.2. During training, listeners performed either semantic or acoustic verification task (accent verification). The "semantic verification" group had to determine whether the sentence was semantically plausible ("true") or not ("false"), and the "acoustic verification" group had to determine whether the talker was "native" or "non-native" Hebrew speaker.

In contrast to previous studies (Banai & Lavner, 2012; Banai & Lavner, 2014) in which participants were given feedback during training, no feedback was provided here because we were concerned that feedback might have different effects on learning in the different conditions .

One of the groups trained on each task (semantic verification and acoustic verification) was exposed to 6 different talkers (high talker variability) during training, while the other was exposed to two talkers only (low talker variability) (see Table 3).

Across groups, the training material consisted a set of 100 sentences, half of which were semantically plausible and the other half were semantically implausible. In the two training protocols with low talker variability, sentences were presented by two talkers: A native Hebrew speaker and a non-native speaker (speaker 1 and speaker 4 respectively, see Table 2). In the training protocols with high talker variability sentences were presented by three native Hebrew speakers (speakers 1-3) and three non-native speakers (speakers 4-6, see Table 2).

Data Analysis

Recognition accuracy for each of the baseline and test conditions was determined for each listener by calculating the percentage of correctly reported words for each sentence heard based on their written transcripts. Only fully and perfectly transcribed words were counted as correct. The only exception was that homophonic spelling errors were not counted because we were interested in the aural recognition of our participants and not in their literacy skills.

Learning and transfer effects were determined based on recognition accuracy in the two test phases. This approach is valid only if group differences in test performance cannot be attributed to initial differences in time-compressed speech recognition across groups (see Banai & Lavner, 2019 for a discussion of why we nevertheless use it). Therefore, data analysis started with a Bayesian analysis comparing the relative likelihoods of two competing hypotheses given baseline data, the null that baseline performance was similar across groups and the alternative that it was not. This analysis was conducted in JASP (version 0.9.2) using a Bayesian ANOVA with group as a fixed effect and participants and items as random effects (JASP Team, 2019). After this analysis suggested that the null hypothesis was more likely than the alternative ($BF < 1/3$; see Results) subsequent analyses focused on performance in the two test phases.

To assess learning, acoustic and semantic transfer we used linear mixed-effects models implemented with the `lmer` function from the `lme4` package in R (Bates et al., 2019). For each test condition, a model with group as a fixed effect and participants and items as random effects was constructed. When these provided evidence for a training effect (that is more accurate performance in trained than in untrained participants), subsequent analyses focused on the potential effects of task and number of talkers, again with participants and items as random effects to account for the variability among participants within a group and across different test items. To account for potential alpha inflation because of repeating the analysis with a subset of the groups, as well as the number of different dependent variables in each test session, critical p-values were determined with a Bonferroni correction, yielding a critical value of $p = 0.0042$. Finally, effects were further explored with a Bayesian ANOVA in JASP with training task and number of talkers as fixed factors and participant and item as random effects.

To analyze the data from the training phase, verification thresholds were calculated for each block of 60 trials as the mean compression level from the last five reversals on that block. Reversals are trials on which the direction of the adaptive track changes. Because the level of compression changed adaptively (2-down/1-up staircase) during training, calculated thresholds reflect the 71% correct on the psychometric function (Levitt, 1971) linking level of compression and verification accuracy. The slopes of the resulting five-block learning curves were calculated, negative slopes are an indication of training-induced learning.

Results

Baseline

Initial performance was similar across groups (see Figure 1; $B < 0.004$, $STE = 0.01$, $t(78) = -0.393$, $p = 0.695$, $BF_{10} = .037$). Given that we failed to reject the null hypothesis that initial performance is similar across groups and the Bayesian analysis further suggested that this null hypothesis is more likely than the alternative, subsequent analyses were conducted on test performance scores (rather than on difference scores) with no further consideration of baseline performance.

Figure 1. Baseline recognition of time-compressed speech (means and 95% CIs)

Immediate Test

Learning. As shown in the leftmost panel of Figure 2A, training led to substantial learning of the trained tokens relative to untrained controls ($B = 0.041$, $STE = 0.009$, $t(78) = 4.42$, $p < 0.0000318$, $BF_{10} = 4993$). Effects sizes (Cohen's d compared to untrained controls) were intermediate ($d = 0.57$) with low-variability acoustic training, and large with the other three training protocols ($d = 1.2$, 1.3 and 1.5 for high-variability acoustic, low-variability semantic and high-variability semantic training, respectively).

To determine if the specific training task or the specific amount of talker variability experienced in training, had an effect on learning, a linear mixed model was constructed with training task, talker variability and their interaction as fixed effects, and item and participant as random effects. Consistent with Figure 2A (left panel), this analysis suggested that low talker variability had a negative influence of approximately 8% on performance in the acoustic task ($B = -0.085$, $STE = 0.036$, $t(60) = -2.266$, $p = 0.021$), although this effect was not significant after accounting for multiple comparisons (given the number of conditions and tests, the Bonferroni corrected alpha is 0.0042). The Bayesian analysis yielded an inconclusive finding ($BF_{10} = 0.84$). The effects of training task ($B = 0.04$, $STE = 0.035$, $t(60) = 1.21$, $p = 0.23$) and the task by variability interaction ($B = 0.06$, $STE = 0.051$, $t(60) = 1.37$, $p = 0.18$) were not significant. The Bayes Factor for the full model was again inconclusive ($BF_{10} = 2.61$). Inconclusive Bayesian outcomes could suggest that the design lacks power to reach a conclusion in favor of either the null or the alternative hypotheses. However, because similar designs and sample sizes yielded conclusive findings in our previous studies (e.g., Banai & Lavner, 2019), we interpret our current failure to obtain conclusive evidence for either the null or the alternative hypotheses to suggest that even if task or talker variability effects exist, they are likely to small to be of interest after a single training session.

Figure 2. Learning and transfer as a function of group. A. Learning and semantic transfer. B. Acoustic transfer. For the trained groups, means and 95% confidence intervals are shown. The grey rectangles mark the 95% confidence intervals of the untrained control group.

Semantic transfer. A visual inspection of Figure 2A shows that performance of the trained groups on the semantic transfer conditions was not noticeably better than that of untrained controls. Therefore, it seems that training failed to yield either semantic transfer (estimated with new sentences recorded by the previously encountered talker,

mid panel of Figure 2A) or acoustic-semantic transfer (estimated with new sentences recorded by a new talker, rightmost panel of Figure 2A). Linear mixed modelling likewise yielded an insignificant group effect (Semantic transfer: $B = 0.001$, $SE = 0.016$, $t(78) = 0.058$, $p = 0.095$; Acoustic-semantic transfer: $B = 0.009$, $SE = 0.01$, $t(78) = 0.99$, $p = 0.324$). The Bayesian analysis suggests that the current data favors the null hypothesis that all groups performed the semantic transfer conditions similarly (semantic transfer: $BF_{10} = 0.028$; acoustic-semantic transfer: $BF_{10} = 0.034$).

Acoustic transfer. In contrast to the lack of semantic transfer as a function of training, training appears to have induced substantial acoustic transfer, as estimated with previously encountered sentences produced by either a new native (Figure 2B, left) or a new nonnative speaker (Figure 2B, right). Linear-mixed modelling confirmed that the effect of groups was significant (native talker: $B = 0.086$, $SE = 0.012$, $t(78) = 7.44$, $p < 0.0001$, $BF_{10} > 10000$; nonnative talker: $B = 0.063$, $SE = 0.009$, $t(78) = 6.92$, $p < 0.0001$, $BF_{10} > 10000$). Nevertheless, the magnitude of the effect compared to untrained controls varied by group (native speaker: Cohen's $d = 1.1, 1.3, 2, 2.3$; nonnative speaker: $d = 0.9, 1.5, 2.3, 2.5$ for the low-variability acoustic, high-variability acoustic, low-variability semantic and high-variability semantic training groups, respectively).

To decipher the contributions of the training task, talker variability and their interaction, a mixed model was fitted with participants and items as random effects. This analysis suggests that semantic training facilitated transfer to both native and nonnative talkers ($B = 0.167$, $SE = 0.044$, $t(60) = 3.792$, $p = 0.0003$; $B = 0.116$, $SE = 0.035$, $t(60) = 3.301$, $p = 0.0016$, respectively). The Bayesian effects were consistent with those of the hypothesis testing ($BF_{10} = 3647$ and 2783 for the native and nonnative talkers, respectively). Talker variability (native talker: $B = -0.02$, $SD = 0.045$, $t(60) = -$

0.448, $p = 0.656$; nonnative talker: $B = -0.065$, $SD = 0.036$, $t(60) = -1.832$; $p = 0.072$) and the task by variability interaction (native talker: $B = -0.019$, $SE = 0.064$, $t(60) = -0.303$, $p = 0.763$; nonnative talker: $B = 0.032$, $SD = 0.050$, $t(60) = 0.64$, $p = 0.522$) had no significant effect. Bayesian Factors for the talker effects were also small ($BF_{10} = 0.327$ and 0.627 for the native and nonnative talkers, respectively).

Delayed Test

Learning. Figure 3A suggests that learning was retained two weeks after training trained groups ($B = 0.031$, $STE = 0.009$, $t(73) = 3.48$, $p = 0.000847$, $BF_{10} = 4.78$). Relative to untrained controls, effect sizes were intermediate in the two acoustic training groups (Cohen's $d = 0.58$, 0.54 in the low and high variability groups respectively), and large in the two semantic training groups (1.2 and 1.3 in the low and high variability groups). According to Figure 3A, performance was better in the semantic than in the acoustic groups, but this difference was not statistically significant in the follow up statistical analysis. A linear mixed model with training task, talker variability and their interaction as fixed effects, and item and participant as random effects yielded no significant effects (task: $B = 0.067$, $STE = 0.040$, $t(56) = 1.68$, $p = 0.0979$; talker variability: $B = -0.010$, $STE = 0.040$, $t(56) = -0.25$, $p = 0.800$; interaction: $B = -0.015$, $STE = 0.056$, $t(56) = -0.27$, $p = 0.786$). The Bayesian analysis was consistent with the conclusion that neither of the fixed effects or their interaction influenced learning ($BF_{10} = 0.063$).

Semantic transfer. Similar to the immediate test, it seems that two weeks after training, the trained groups performed within the range of the control group on the two semantic transfer tests (Figure 3A, mid and rightmost sections). Linear mixed modelling likewise yielded an insignificant group effect (Semantic transfer: $B = 0.009$, $SE = 0.010$, $t(73) = 0.908$, $p = 0.367$; Acoustic-semantic transfer: $B = 0.011$, $SE = 0.01$, $t(73) = 1.14$, $p =$

0.259). The Bayesian analysis likewise suggests that the current data favors the null hypothesis that all groups performed the semantic transfer conditions similarly (semantic transfer: $BF_{10} = 0.036$; acoustic-semantic transfer: $BF_{10} = 0$).

Acoustic transfer. In contrast to the lack of semantic transfer as a function of training, training appears to have induced substantial acoustic transfer, as estimated with previously encountered sentences produced by either a new native (Figure 3B, left) or a new nonnative speaker (Figure 3B, right). The linear mixed model did not converge for the native talker, and the Bayesian analysis was inconclusive ($BF_{10} = 0.383$). For the nonnative talker, linear-mixed modelling confirmed that the effect of groups was significant ($B = 0.033$, $SE = 0.010$, $t(73) = 3.17$, $p = 0.0022$, $BF_{10} = 1.27$). Nevertheless, the magnitude of the effect compared to untrained controls were overall smaller in the delayed test (native speaker: Cohen's $d = 0.5, 0.1, 0.6, 0.8$; nonnative speaker: $d = 0.4, 0.6, 1.0, 1.1$ for the low-variability acoustic, high-variability acoustic, low-variability semantic and high-variability semantic training groups, respectively). As for learning, the follow up analysis yielded no significant effects. Semantic training may have facilitated performance to some extent, but the effect was not significant (native talker: $B = 0.11$, $STE = 0.055$, $t(56) = 1.94$, $p = 0.057$; nonnative talker: $B = 0.066$, $STE = 0.048$, $t(56) = 1.37$, $p = 0.175$). Likewise, the effects of talker variability and the task by talker interaction were insignificant. This was true for the native (variability effect: $B = 0.073$, $STE = 0.553$, $t(56) = 1.32$, $p = 0.191$; interaction effect: $B = -0.091$, $STE = 0.0782$, $t = -1.16$, $p = 0.251$) as well as for the nonnative (variability: $B = -0.0173$, $STE = 0.048$, $t(56) = -0.36$, $p = 0.721$; interaction: $B = 0.013$, $STE = 0.068$, $t(56) = 0.19$, $p = 0.850$) talker. The Bayesian analyses were likewise consistent with the conclusion that neither of the effects or their interaction influenced acoustic transfer ($BF_{10} = 0.070$ and 0.072 for the whole model, for the native and nonnative talkers, respectively).

Figure 3. Performance in the delayed test. A. Learning and semantic transfer. B. Acoustic transfer. For each group, means and 95% CIs are shown.

For details see Figure 2.

Training

Sixty-four participants participated in the training phase. The training data of three participants were lost due to technical issues. Therefore, this section presents the data of 61 participants, 16 of the training group that trained on the high variability-semantic verification task, and 15 of each of the other training groups.

Acoustic verification task. The group learning curves (Figure 4, left panel) suggest that mean verification thresholds improved during training, although the improvement is hard to observe in the low-variability (2-talkers) group, which reached very low compression levels by the first training block (the lowest allowed in our adaptive procedure). In the two-talker group, 11/15 individual curves were negative. The remaining four participants reached maximal compression (0.2) on the first block and remained in this level throughout training. The mean slope was significantly negative, as shown in Table 4. In the six-talkers training group, 14/15 individual curves were negative and the mean slope was also significantly negative (see Table 4).

Figure 4. Learning during acoustic (left panel) and semantic (right panel) training. Mean verification thresholds and 95% confidence intervals across the five training blocks are shown for each group.

As seen in Figure 4, the low-variability training group achieved low compression threshold (nearly 0.2 of the uncompressed duration) from the onset of training. However, the high-variability group showed gradual improvement. The mean slope of

the learning curve was significantly steeper in the high variability than in the low variability training group ($t(14.79) = -3.226$, $p < .006$).

Semantic verification task. The group learning curves (Figure 4, right panel) show that mean verification thresholds improved during training. 12 individual curves (out of 15 in the two-talkers and out of 16 in the six-talker group) were negative. The mean slopes of the learning curves in the two semantic groups were significantly negative (Table 4) but didn't differ from each other ($t(29) = .857$, $p = .399$).

Table 4. Mean learning curve slopes and confidence intervals

Group	Mean slope ($\pm$ SD)	95% confidence interval
low variability- accent	-0.0024 ± 0.0038	-0.00444 -0.00029
high variability- accent	-0.0212 ± 0.0223	-0.0335 -0.0088
low variability- semantic	-0.0120 ± 0.0096	-0.0173 -0.0067
high variability- semantic	-0.0081 ± 0.0151	-0.0161 -0.00008

Discussion

The current study aimed to investigate the perceptual learning of time-compressed speech following training with different tasks and multiple talkers. As in previous studies, the perceptual learning of time-compressed speech was robust, but the transfer of learning to novel stimuli (semantic transfer) was limited. Although learning and acoustic transfer occurred with all protocols, practice on semantic verification yielded more learning and acoustic transfer than practice on accent verification. In contrast to previous studies with other forms of acoustically difficult speech, we find no effect of talker variability on the transfer of learning to new stimuli. Improvements were generally maintained between the immediate and delayed tests, but additional

improvements among untrained participants make outcomes from this measurement hard to discuss. The following discussion is therefore focused on the findings from the immediate test.

The current findings are partially consistent with our first hypothesis *that all types of training should have resulted in both learning and transfer*. Compared to untrained controls, training resulted in learning as well as in acoustic transfer (to native and non-native speakers) in all trained groups (see Figure 2). On the other hand, neither semantic nor acoustic-semantic transfer were observed. These findings are consistent with previous findings that semantic transfer requires more than a single training session to emerge (Banai & Lavner, 2014, 2019), and fail to replicate the small semantic transfer effects observed in previous single-training-session studies (Gabay, Karni, & Banai, 2017; Manheim et al., 2018). One way to reconcile those differences is that the greater number of test conditions included in the current study was sufficient to induce some rapid learning effects in the untrained group, which masked the slight contribution of additional practice. Inspection of the confidence intervals of the control group in the baseline phase (Figure 1) and in the learning condition during the test phase (Figure 2), certainly suggest some learning even without training. Additionally, trained listeners in the current study had fewer opportunities to consolidate and reconsolidate what they have learned, also potentially reducing transfer (Banai & Lavner, 2019; Wright Wilson & Sabin, 2010; Xie, Earle & Myers, 2018). Another factor that may have limited semantic transfer in the current study is the exposure to non-native speakers during the training phase. Exposure to non-native talkers, posed a ‘double challenge’ to our participants on half of the training trials (having to deal with both accented and time-compressed speech), and may have thus limited transfer (Wade et al., 2007). Either

way, semantic transfer following training on time-compressed speech appears limited, and the task and talker manipulations in the current study failed to broaden it.

We also hypothesized that *training on the semantic task should yield more learning and transfer than training on the acoustic task*. On the conditions in which training effects were observed (i.e., learning and acoustic transfer), current findings are consistent with this hypothesis on both the immediate (Figure 2) and delayed (Figure 3) tests. The semantic training groups generally outperformed the acoustic training groups with larger effect sizes relative to untrained controls, although effect sizes were smaller in the delayed test, probably due to further improvements among untrained participants due to repeated exposures to time-compressed speech. As suggested in the introduction, we attribute this task effect to the greater similarity between the semantic verification task and the recognition task used to assess learning. This finding is consistent with the transfer appropriate processing (TAP) theory of learning and memory, which proposes that performance is maximized when the training and test tasks are similar to each other (Morris et al., 1977), as well as with previous findings on the task specificity of perceptual learning (reviewed in the introduction section).

It could be argued that rather than similarity to the test tasks, the accent verification task used for acoustic training was easier than the semantic-verification task, and therefore did not sufficiently challenge listeners during training. We do not think this is the case, because although performance on the accent task was very good from the onset of training with two talkers, with six talkers both initial performance during the training phase and the shape of the learning curve appear similar to those observed with the semantic task (Figure 4). We note that the accent verification task was selected after considering several options for the acoustic task (e.g., gender and intonation verification). To keep the characteristics of the adaptive training procedure

similar across the semantic and acoustic tasks, we decided to use a binary decision task similar to that used for semantic verification in our previous studies. Thus, listeners in the acoustic groups had to decide whether on each sentence the talker was a native or a non-native speaker of Hebrew. They did not have to identify the non-native accent or the particular talker. Gender identification was ruled out because a previous study with vocoded speech reported a ceiling effect (Loebach et al., 2008), and a brief pilot suggested that the same is true for highly compressed speech as well. Intonation identification (question vs. statement) was also piloted and deemed too easy. Accent verification was thus selected as the most challenging of the options we considered.

Lastly, our findings provided no support for our final hypothesis, that *talker variability should interrupt learning with either acoustic or semantic training but facilitate the transfer of learning following semantic training*. Although the learning profiles in the acoustic task (Figure 4) could perhaps suggest faster learning (within the first block of practice) with fewer talkers, test phase performance in the current study was not significantly influenced by talker variability. This is in contrast to previous findings (reviewed in the Introduction) that talker variability improves the transfer of learning to new speech materials (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Brosseau-Lapr et al., 2013; Clopper & Pisoni, 2004; Lively et al., 1993; Sidaras et al., 2009). It could be the case that different acoustic features respond differently to talker variability. Talker variability creates different realizations of new phonetic categories and non-native speech (produced by speakers sharing an accent), and may thus support the learning of category or accent features that are shared across talkers, as reported in previous studies (Baese-Berk et al., 2013; Bradlow et al., 1999; Bradlow & Bent, 2008; Brosseau-Lapr et al., 2013; Lively et al., 1993; Lively et al., 1994). On the other hand, the uniform time compression we used here has a significant effect on quasi-periodic

elements such as vowels, which are similar across talkers. Therefore, time compression may eliminate some of the systemic variation in the realization of vowels which has been suggested to contribute to talker variability effects (Brosseau-Lapr et al., 2013; Sidaras et al., 2009). Furthermore, although average pre-compression speech rates of the talkers in the high-variability protocol were quite variable (92-136 words/min), within talker rate variability was also substantial (see Table 2). This means that during training listeners would have encountered rapid tokens produced by talkers with slower mean rates as well as rapid tokens produced by talkers with higher mean rates, making the talkers less distinct.

Due to the limitations on the semantic and acoustic-semantic transfer of time-compressed speech, which were already discussed, we are reluctant to offer further discussion of the null effect of talker variability. Nevertheless, we note that even in the cases of category and accent learning talker variability is not always helpful (see Bent & Frush Holt, 2013; McLennan & Luce, 2005; Wade et al., 2007). Furthermore, recent studies call for a more nuanced approach to talker variability, especially in the case of non-native speakers (Baese-Berk & Morrill, 2019; Vaughn, Baese-Berk & Idemaru, 2019).

References

- Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America*, 126(5), 2649-2659.
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1515), 285-299.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, 133(3), EL174-EL180.

Baese-Berk, M. M., & Morrill, T. H. (2019). Perceptual Consequences of Variability in Native and Non-Native Speech. *Phonetica*, 76(2-3), 126-141.

Banai, K., & Lavner, Y. (2012). Perceptual learning of time-compressed speech: More than rapid adaptation. *PloS one*, 7(10), e47099.

Banai, K., & Lavner, Y. (2014). The effects of training length on the perceptual learning of time-compressed speech and its generalization. *The Journal of the Acoustical Society of America*, 136(4), 1908-1917.

Banai, K., & Lavner, Y. (2016). The effects of exposure and training on the perception of time-compressed speech in native versus nonnative listeners. *The Journal of the Acoustical Society of America*, 140(3), 1686-1696.

Banai, K., & Lavner, Y. (2019). Effects of stimulus repetition and training schedule on the perceptual learning of time-compressed speech and its transfer. *Attention, Perception, & Psychophysics*, 1-12.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christiansen, R., Singmann, H., Dai B., Scheipl, F., & Grothendieck, G. (2019). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-21, <http://CRAN.R-project.org/package=lme4>.

Bent, T., & Frush Holt, R. (2013). The influence of talker and foreign-accent variability on spoken word identification. *The Journal of the Acoustical Society of America*, 133(3), 1677-1686.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & psychophysics*, 61(5), 977-985.

- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.
- Brosseau-Lapr, F., Rvachew, S., Clayards, M., & Dickson, D. (2013). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics*, 34(3), 419-441.
- Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, 47(3), 207-238.
- Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of experimental psychology: Human perception and performance*, 36(3), 704.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2), 222.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 914.
- Gabay, Y., Karni, A., & Banai, K. (2017). The perceptual learning of time-compressed speech: A comparison of training protocols with different levels of difficulty. *PloS one*, 12(5), e0176488.
- Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), 152.
- Green, C. S., Banai, K., Lu, Z. L., & Bavelier, D. (2018). Perceptual learning. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, 2, 1-47.

Halliday, L. F., Moore, D. R., Taylor, J. L., & Amitay, S. (2011). Dimension-specific attention directs learning and listening on auditory training tasks. *Attention, Perception, & Psychophysics*, 73(5), 1329-1335.

Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature neuroscience*, 7(10), 1055.

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1), 283.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive psychology*, 51(2), 141-178.

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2B), 467-477.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. I., & Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. *The Journal of the acoustical society of America*, 96(4), 2076-2087.

Loebach, J. L., Bent, T., & Pisoni, D. B. (2008). Multiple routes to the perceptual learning of speech. *The Journal of the Acoustical Society of America*, 124(1), 552-561.

Manheim, M., Lavie, L., & Banai, K. (2018). Age, Hearing, and the Perceptual Learning of Rapid Speech. *Trends in hearing*, 22, 2331216518778651.

- Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), 676.
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 306.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519-533.
- Nahum, M., Nelken, I., & Ahissar, M. (2010). Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning. *Vision research*, 50(4), 391-401.
- Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *Journal of neuroscience*, 26(18), 4970-4982.
- Prior, A., & Bentin, S. (2006). Differential integration efforts of mandatory and optional sentence constituents. *Psychophysiology*, 43(5), 440-449.
- Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 539.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207-1218.
- Sidasas, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306-3316.

Simhony, M. A., Grinberg, M., Lavie, L., & Banai, K. (2014). Rapid adaptation to time-compressed speech in young and older adults. *Journal of basic and clinical physiology and pharmacology*, 25(3), 285-288.

Vaughn, C., Baese-Berk, M., & Idemaru, K. (2019). Re-examining phonetic variability in native and non-native speech. *Phonetica*, 76(5), 327-358.

Verhelst, W., & Roelands, M. (1993). An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. In *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 2, pp. 554-557).

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica*, 64(2-3), 122-144.

Wright, B. A., Wilson, R. M., & Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task. *Journal of Neuroscience*, 30(35), 11635-11639.

Xie, X., Earle, F. S., & Myers, E. B. (2018). Sleep facilitates generalisation of accent adaptation to a new talker. *Language, cognition and neuroscience*, 33(2), 196-210.