Attention, Perception, & Psychophysics (2019) 81:2944–2955
https://doi.org/10.3758/s13414-019-01714-7

Effects of stimulus repetition and training schedule on the perceptual learning of time-compressed speech and its transfer

Karen Banai[1] · Yizhar Lavner[2]

Published online: 3 June 2019

Abstract
Perceptual learning can facilitate the recognition of hard-to-perceive (e.g., time-compressed or spectrally-degraded) speech. Although the learning induced by training with time-compressed speech is robust, previous findings suggest that intensive training yields learning that is partially specific to the items encountered during practice. Here, we asked whether three parameters of the training procedure – the overall number of training trials (training intensity), how these trials are distributed across sessions, and the number of semantically different items encountered during training (set size) – influence learning and transfer. Different groups of participants (69 normal-hearing young adults; nine to 11 participants/group) completed different training protocols (or served as an untrained control group) and tested on the recognition of time-compressed sentences taken from the training set (learning), new time-compressed sentences presented by the trained talker (semantic transfer), and time-compressed sentences taken from the training set but presented by a different talker (acoustic transfer). Compared to untrained listeners, all training protocols yielded both learning and transfer. More intense training resulted in greater item-specific learning and greater acoustic transfer than less intense training with the same number of training sessions. Training on a smaller set size (i.e., greater token repetition during training) also resulted in greater acoustic transfer, whereas distributing practice over a number of sessions improved semantic transfer. Together, these data suggest that whereas practice on a small set that results in stimulus repetition during training is not harmful for learning, distributed training can support transfer to new stimuli, perhaps because it provides multiple opportunities to consolidate learning.

Keywords Degraded speech · Rapid speech · Auditory learning · Generalization · Time-compressed speech

Introduction

The perception of both naturally (e.g., rapid, dysarthric) and artificially (e.g., time-compressed, noise-vocoded) degraded speech improves substantially via rapid perceptual learning (Borrie, Lansford, & Barrett, 2017; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; for review see Samuel & Kraljic, 2009). Furthermore, this rapid learning often leads to the transfer of learning to new materials, which depends on the acoustic or acoustic-phonetic similarities of previously experienced and newly encountered items (Adank & Janse, 2009; Borrie et al., 2017; Bradlow & Bent, 2008; Dupoux & Green, 1997; Huyck, Smith, Hawkins, & Johnsrude, 2017; Loebach, Pisoni, & Svirsky, 2009; Peelle & Wingfield, 2005). On the other hand, attempts to harness this remarkable plasticity for hearing rehabilitation through longer-term training programs were often disappointing because improvements were specific to the trained materials (Henshaw & Ferguson, 2013; Saunders et al., 2016). Suggesting that this specificity is not fully related to the characteristics of the trained populations (e.g., older adults or hearing-impaired individuals), intensive training on time-compressed speech resulted in learning that was more specific than that observed following rapid learning even in young, normal-hearing listeners. For example, although brief practice with time-compressed speech resulted in improved processing of new natural-fast speech tokens (Adank & Janse, 2009), no such transfer was observed with more intensive practice (Manheim, Lavie, & Banai, 2018). We now ask whether for time-compressed speech, procedural aspects associated with the learning experience influence the degree of learning and specificity. Specifically, we examined the effects of three

✉ Karen Banai
kbanai@research.haifa.ac.il

[1] Department of Communication Sciences and Disorders, University of Haifa, Mt. Carmel, 3498838 Haifa, Israel

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

training-related factors (the number of training sessions, the size of the training set, and the number of trials in each session) on the learning of time-compressed speech and its generalization to new tokens and to a new talker.

The perceptual learning that emerges following intensive training with time-compressed speech is robust (with Cohen effect sizes > 0.8 compared to untrained controls), but fairly specific. In a series of previous studies, participants were trained with sets of 100 semantically different sentences, all recorded by the same talker, and presented as time-compressed speech. Each sentence was presented several times during training (Banai & Lavner, 2014, 2016). Learning and transfer were tested by comparing the performance of trained and untrained participants on a series of tests in which stimuli were either taken from the training set (to assess learning), or were somehow different (to assess transfer). Transfer conditions differed from the trained condition either acoustically (a novel talker was used), semantically (new sentences were used) or both acoustically and semantically. Trained participants recognized trained sentences, new sentences, and trained sentences with new acoustics more accurately than untrained participants, suggesting both semantic and acoustic transfer of learning. On the other hand, no transfer was found when the test materials differed from the trained materials both acoustically and semantically (Manheim et al., 2018). This pattern of findings is consistent with the Reverse Hierarchy Theory of perceptual learning (RHT, Ahissar, Nahum, Nelken, & Hochstein, 2009). According to the RHT, by default, speech recognition is based on high-level representations, which do not contain detailed information about the fine-grained acoustic-phonetic structure of the item they represent. In the context of speech, this suggests that when listeners attempt to recognize speech, they are tuned to the content and therefore rely on lexical or semantic representations rather than on low-level phonological or acoustic-phonetic ones. On the other hand, if listening for details (e.g., when a minimal pair discrimination is required in a situation with limited lexical information), listeners can rely on acoustic-phonetic cues (Nahum, Nelken & Ahissar, 2008). Novel acoustically degraded speech does not match available high-level representations, making its accurate recognition dependent on the use of lower-level non-default representations that naïve listeners do not automatically access. Data from previous studies suggest that intensive training makes these low-level representations more usable (Francis, Nusbaum, & Fenn, 2007; Nahum, Nelken, & Ahissar, 2010). However, as low-level representations are activated by the stimuli encountered in training, transfer is expected to be limited to novel stimuli that share the same acoustic-phonetic structure of the trained stimuli, explaining why the transfer of training-induced learning of time-compressed speech is incomplete.

While learning specificity might be an inherent (Fiorentini & Berardi, 1980; Hussain, McGraw, Sekuler, & Bennett, 2012), or even desirable, feature of perceptual systems, the conditions modulating this specificity in the case of distorted speech recognition are not well understood. Specifying and characterizing these conditions is important both for the understanding of the perceptual learning that might occur with changing acoustic circumstances, such as when a hearing impaired individual starts using a hearing aid or a cochlear implant and for the potential use of structured training protocols in education (e.g., second language learning) and rehabilitation. The use of low-level speech representations is key for the recognition of perceptually difficult speech (Ahissar et al., 2009; Mattys, Brooks, & Cooke, 2009). According to the RHT, for training-induced learning to transfer to new tokens with shared acoustic properties, the training protocol has to "teach" listeners to rely on low-level speech representations instead of the default reliance on high-level cues. The literature on the perceptual learning of speech suggests that a few factors associated with the training experience might support such shift. One such factor is the use of adaptive training. Although perceptual learning for speech occurs with different forms of training, a few studies suggest that adaptive training protocols in which the level of speech distortion increases gradually with training supports learning (Gabay, Karni, & Banai, 2017; Svirsky, Talavage, Sinha, Neuburger, & Azadpour, 2015), consistent with RHT predictions. For example, in comparisons between adaptive and non-adaptive training protocols, adaptive training facilitated the learning of time-compressed speech and its acoustic transfer (Gabay et al., 2017). Adaptive training also facilitated the transfer of learning of spectrally degraded speech (Svirsky et al., 2015). Therefore, an adaptive training procedure was used for all training protocols in the current study. Other factors that may support learning and transfer – the overall number of training trials (training length) (Banai & Lavner, 2014; Nahum et al., 2010), the opportunity to consolidate learning over multiple training sessions (Fenn, Nusbaum, & Margoliash, 2003), and the number of different items encountered in training (training-set size) (Greenspan, Nusbaum, & Pisoni, 1988; Lively, Logan, & Pisoni, 1993) – are discussed in the following paragraphs.

Multi-session speech training studies suggest that the perceptual learning of speech is often not exhausted with brief exposure, but rather continues across hundreds or more trials (Banai & Lavner, 2012, 2014). These studies suggest that more training leads to more learning, but the effects of extended training on the transfer of learning to untrained tokens are inconsistent. On the one hand, the learning of specific phonetic discriminations is often quite specific to the trained contrast (Lively et al., 1993; Nahum et al., 2010). These studies suggest that listeners can learn to use the relevant low-level cues required for the discrimination, but this learning is fairly specific. On the other hand, when speech in noise, time-compressed speech, or spectrally-degraded speech were trained with sentence length materials, learning continued across sessions even though specific sentences were not repeated during training or between training and test (Karawani, Bitan, Attias, & Banai, 2015; Song, Skoe, Banai, & Kraus, 2012; Svirsky et al., 2015). Although multiple factors might account for the difference in specificity between the discrimination study of Nahum et al. and the other studies, the fact that in the latter studies training was not focused on a specific contrast may have resulted in greater generalization. In a study of time-compressed speech learning, listeners practiced with an adaptive training protocol for either one or three sessions (Banai & Lavner, 2014). Learning of the trained tokens, semantic transfer and acoustic transfer were all greater in the three-session group than in the single-session training group. However, in the three-session protocol listeners received more training sessions as well as an overall larger number of training trials. Thus it is not clear whether the three-session group benefit resulted from the overall larger number of learning opportunities or from the spacing of learning opportunities across sessions, which has been amply documented for the learning of verbal information (for review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Therefore, and although a recent study found no spacing effect for the perceptual learning of speech in hearing-aid users (Tye-Murray, Spehar, Barcroft, & Sommers, 2017), in the present study we manipulated both the overall number of trials and their spacing across sessions.

In addition to the number of training trials, another factor reported to widen the scope of acoustic-semantic transfer is the variability of the training set. Specifically, for learning a non-native speech contrast (Lively et al., 1993) as well as for improving the recognition of accented speech (Bradlow & Bent, 2008), it has been reported that experience with a larger number of different talkers during practice supports the transfer of learning to new talkers ("acoustic transfer"). A common interpretation for these findings is that greater stimulus variability during exposure facilitates adaptation to talker-independent characteristics of the accented or non-native speech (Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008). However, in the case of time-compressed speech, the pattern of learning that emerges following intensive training shows some degree of specificity to the trained tokens, even without the introduction of a new talker (Banai & Lavner, 2014; Manheim et al., 2018). This suggests that our training protocols resulted in listeners focusing on high-level cues rather than on the low-level cues, which would have afforded generalization to new tokens that share the acoustic-phonetic structure (i.e., the same talker) of the trained stimuli (Nahum, Nelken, & Ahissar, 2008). We reasoned that a "denser sampling" of the relevant acoustic-phonetic space (by increasing the number of different tokens) should support the transfer of learning to new tokens taken from the same acoustic space (Greenspan et al., 1988). Therefore, instead of manipulating the number of trained talkers, we manipulated the number of different tokens that were introduced by a single talker during training. We asked whether the number of different tokens had any effect of transfer to new semantics (new sentences presented by a familiar talker) and to transfer to new acoustics (familiar sentences presented by a new talker).

Methods

Participants

Seventy undergraduate University of Haifa students (mean age = 25 years, SD = 2; 67 females) with no prior experience with time-compressed speech participated in this study after giving their consents in accordance with the guidelines of the Faculty of Social Welfare and Health Sciences ethics committee (permit 199/12). By self-report all participants were native speakers of Hebrew with normal hearing and had no history of learning or attention deficits. One participant dropped out and did not complete all assessments. Therefore, we report data from 69 participants (nine to 11 per group, see Table 1). Sample size was determined based on previous studies that showed that groups of ten participants were sufficient to uncover learning and generalization in both single and multi-session training protocols with intermediate to large effect sizes (Banai & Lavner, 2014, 2016). Therefore, this number should allow for the detection of any facilitative effects of the manipulations used in the current study.

Stimuli

A set of 260 sentences in Hebrew, five to six words each (based on Prior & Bentin, 2006), were used in this study. Half of the sentences were semantically plausible (e.g., "the grumpy chef prepared a great meal") while the other half were implausible (e.g., "the woolly sheep measured the green dress"). All sentences were recorded and sampled at 44 kHz by two young male native Hebrew speakers using a standard microphone and PC sound card. One of the speakers, with a natural speech rate of 107 words/min and average fundamental frequency (F0) of 106 Hz (range: 86–132 Hz, mean formant dispersion: 929 Hz; F1: M = 601 Hz, SD = 77, F2: M = 1,225 Hz, SD = 118, F3: M = 2,415 Hz, SD = 199), recorded all 260 sentences. This speaker was designated "the familiar talker," and his recordings were used for both training and testing. The other speaker (the new talker), with a natural speech rate of 124 words/min and average F0 of 108 Hz (range: 81–145 Hz, mean formant dispersion: 1,008 Hz; F1: M = 587 Hz, SD = 78, F2: M = 1,282 Hz, SD = 113, F3: M = 2,604 Hz, SD = 218) recorded 20 of the sentences that were used to test for cross-talker transfer of learning. Although the use of a male familiar talker and a female transfer talker would have provided a stronger test of acoustic transfer, we note that the two male talkers were clearly distinct, and in addition to the differences in formant values and dispersion shown above, give the impression of having different voice qualities. In a

Table 1 Training groups and protocols

	Spacing (number of training sessions)			
	One (massed training)		Four (distributed training)	
Training intensity[a] Training set size[b]	60 trials/session	240 trials/session	60 trials/session	240 trials/session
Large (240 different sentences)	--[c]	L240X1 (n = 10)	L60X4 (n = 9)	L240X4 (n = 11)
Small (60 different sentences)	--[d]	S240X1 (n = 9)	S60X4 (n = 10)	S240X4 (n = 10)

Each cell shows the acronym for one of the trained groups based on the three experimental variables: training intensity (60 or 240 trials/session), training set size (large or small) and spacing. The number of participants per group is given in parentheses within each cell

[a] The length of each of the training sessions. A greater number of trials is considered more intense training

[b] The number of different sentences that were included in the training set for a given protocol; L = a large training set of 240 different sentences; S = a small training set of 60 different sentences; 1 = a single training session; 4 = 4 training sessions

[c] Impossible to implement

[d] Deemed unnecessary based on previous findings

previous study with these two speakers and an additional female speaker, transfer across the two male talkers was similar in magnitude to the transfer across genders (Banai & Lavner, 2012). A WSOLA algorithm (Verhelst & Roelands, 1993) was used to time-compress the sentences in the time domain.

Sentences were used as follows: 240 sentences recorded by the familiar talker were used for training in the large stimulus-set conditions (see below). Sixty of these sentences (randomly selected) were included in the small-set training conditions. Twenty of these 60 sentences were used to assess baseline performance as well as training-induced learning. The remaining 20 sentences from the familiar talker (which were not included in any other set) were used to assess transfer to new items (semantic or cross-token transfer). The 20 sentences recorded by the new talker were assigned to the cross-talker (acoustic) transfer set.

Overview of the design, experiment schedule, and test conditions

Participants were randomly assigned to one of seven groups – an untrained control group (C, n = 10) and six training groups. Training focused on the recognition of time-compressed speech using different training protocols (described below). All listeners participated in an initial baseline assessment and a test session conducted 14–21 days apart. Listeners assigned to the training groups completed additional training in between the baseline and test sessions, as described below.

Phases and tasks

Baseline To assess initial performance with time-compressed speech, 20 sentences recorded by the familiar talker were presented compressed to 30% of their naturally spoken duration. This rate was used in our previous studies (Banai & Lavner, 2014, 2016; Gabay et al., 2017), and was selected initially based on a pilot study that suggested that for this talker, a compression of 35% yielded very accurate performance, which left little room for training-related improvements. Listeners were instructed to listen to each sentence and write it down as accurately as they could. The task was self-paced and no feedback was provided.

Training An adaptive sentence verification task was used across all six training protocols. For this, blocks of 60 sentences, selected at random (without replacement) from the relevant training set were presented. Each training block always contained 30 plausible and 30 implausible sentences. Randomization was applied as follows: First an individual trial was selected as plausible or implausible, and then a sentence was randomly selected from the appropriate group of sentences within the stimulus set. The process restarted once all stimuli of a given set were presented.

After listening to each sentence listeners were asked to indicate whether it was plausible or not by clicking a corresponding button on the computer screen. Feedback was provided for both correct (a smiling cartoon face) and incorrect (a sad cartoon face) responses. On the first trial of each training block, the sentence was compressed to 63% of its natural rate. Subsequently, the level of compression changed based on participants' responses in a two-down/one-up staircase procedure (Levitt, 1971) to a maximum of 20% of the natural speech rate of the familiar talker. The staircase comprised of 25 logarithmically equal steps such that after two initially correct responses compression dropped to 0.52, then to 0.44 and so forth. All stimuli during training were presented by a single talker (the trained talker, see below). Previous work suggests that this adaptive protocol yields more learning of time-compressed speech than other possible protocols (Gabay et al., 2017).

The six training protocols differed on three parameters (see Table 1): (1) training set size (large stimulus set – 240 different sentence or small stimulus set – 60 different sentences); (2) spacing of the training trials (all massed in a single session or distributed across four sessions, conducted on different days); (3) intensity of training (60 trials per session, delivered within approximately 5–7 min, or 240 trials per session, approximately 25–30 min). The specific protocols were as follows:

1) Massed training with a large stimulus set (L240X1): Participants completed a single training session of 240 different sentences, presented in four blocks of 60 trials each.
2) Massed training with a small stimulus set (S240X1): Participants completed the same training as above with a stimulus set of 60 different sentences, such that each sentence was repeated four times during training.
3) Brief distributed training with a large stimulus set (L60X4): Participants completed four training sessions of 60 trials each. A different set of 60 sentences was presented in each session such that each sentence was encountered only once during training. This protocol is identical to the L240X1 protocol, but distributed over four sessions.
4) Brief distributed training with a small stimulus set (S60X4): Participants completed four training sessions of 60 trials each. The same 60 sentences were presented in each session such that each sentence was presented four times during training. This protocol is identical to the S240X1 protocol, but distributed over four sessions.
5) Long distributed training with a large stimulus set (L240X4): Participants completed 960 training trials distributed over four training sessions of 240 trials each (four blocks of 60 trials in each session). Each of the 240 different sentences was presented once on each training session.
6) Long distributed training with a small stimulus set (S240X4): As in the previous protocol, participants completed 960 training trials distributed over four training sessions of 240 trials each. However as only 60 different sentences were used in training, each sentence was encountered four times per session, and 16 times throughout training.

The different training protocols were designed to test the effects of three training-related variables – training intensity, training distribution (massed vs. spaced) and training set size, but the design is not fully factorial for three reasons: First, the inclusion of one untrained control group would not have fitted a factorial design. Second, according to the findings of a previous study from our lab (Banai & Lavner, 2014), little semantic transfer was observed with a single training session of 100 trials. Therefore, we decided against the inclusion of a training group with a single brief training session. Third, even if such a group would have been included, an exhaustive pairing of the three training variables is impossible because it is not possible to administer a large training set with a single 60-trial training session.

Test Participants were tested on three tests of 20 sentences each, all compressed to 30% of their natural duration and presented at a fixed order (learning, transfer to new items, transfer to new talker). The writing task from the baseline phase was used again. Sentences in the three tests shared different features with the trained stimuli as follows:

1) Learning: This test was designed to assess performance with stimuli that were encountered (by the trained listeners during training). The 20 sentences used for baseline assessment were presented again to all participants. A comparison of performance on this test across groups indicates whether training resulted in improved performance relative to controls who participated in the baseline assessment but received no training and whether learning was influenced by stimulus set size, spacing, and training intensity.
2) Transfer to new items (semantic transfer): To assess performance with new sentences that were not encountered with either of the previous phases, 20 new sentences (presented by the familiar, trained talker) were presented. A comparison of performance on this test across group indicates whether there was transfer of learning to new sentences and whether transfer depended on stimulus set size, spacing, or training intensity.
3) Transfer to new talker (acoustic transfer): The 20 sentences from the baseline and learning tests were presented again by a new talker to assess performance with time-compressed speech with new acoustic features. A comparison of performance on this test across groups indicates whether learning was specific to the acoustics of the familiar talker and whether this specificity was influenced by set size, spacing, or training intensity.

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 on their written transcripts. Errors reflecting erroneous recognition of the auditory words (e.g., errors in tense or suffix) were all counted as errors, but homophonic spelling errors were ignored. Raw recognition scores (percent correct) are reported in the text and figures for ease of interpretation, but statistical analyses were carried out on rationalized arcsine transformed scores (Studebaker, 1985) because raw scores were not normally distributed.

The current study does not include a pre-test of all test conditions. Therefore, determination of learning and transfer effects is based on test-phase performance. Such analysis is only valid to the extent that baseline performance is similar across the different groups, which is why a baseline assessment of a single condition of time-compressed speech was included. Several reasons led to the selection of this perhaps less-than-ideal design. First, multiple studies (e.g., Altmann & Young, 1993; Dupoux & Green, 1997) have documented a very rapid phase of time-compressed speech learning. Substantial rapid learning during the pre-test can "mask" potential differences between the training protocols, especially on the transfer conditions in which training-related effects can be expected to be subtle (see Banai & Lavner, 2014). Second, a similar design with minimal pre-testing seems to dominate studies of the perceptual learning of speech (e.g., Borrie et al., 2017; Bradlow & Bent, 2008; Peelle & Wingfield, 2005). Finally, the major outcomes of strong training-related learning and more subtle training-related semantic transfer were qualitatively similar when compared between previous studies in which difference scores were used (Banai & Lavner, 2012; Manheim et al., 2018) and those in which the design was similar to that of the current study (Banai & Lavner, 2014; Gabay et al., 2017).

Data were analyzed with a one-way analysis of variance (ANOVA, with group as independent factor) of the data from each of the test conditions followed by a Bayesian analysis intended to compare the relative odds of two competing hypotheses (a null hypothesis, H0, of no group differences and an alternative hypothesis, H1, that not all groups perform equally) given the performance data, assuming equal priors. In addition, planned comparisons (t tests) were conducted on the test phase data. These were used to determine whether training had an effect on performance in that condition, and whether the size of the training set, spacing or training intensity had any effect on learning and its transfer to stimuli that are somewhat different from those encountered during training. Cross-token (semantic) and cross-talker (acoustic) transfer were tested. To these ends, six contrasts were calculated (see Table 2). One planned comparison was carried out to test the hypothesis that training had an influence on the recognition of time-compressed speech. Test phase performance was compared between the control group on the one hand and all six training groups on the other. Five additional comparisons were used to test hypotheses about the effects of the three training related manipulations employed in this study – training intensity, the spacing of training across sessions, and training set size, as follows:

1) The effect of spacing was tested by comparing the two groups that practiced for 240 trials massed in a single session (S240X1 and L240X1) to the two groups that practiced for the same number of trials distributed across four training sessions (S60X4, L60X4).
2) The effect of training intensity, that is whether increasing the duration of each practice session and thus the overall number trials encountered during training, was tested by comparing the two groups that received brief training sessions (S60X4 and L60X4) to the two groups that received longer training sessions (S240X4 and L240X4).
3) The effect of the training set size (i.e., whether encountering a greater variety of stimuli during training is favorable) was tested by comparing the groups that practiced with a large training set (L240X1, L60X4 and L240X4) to those that practiced with a small training set (S240X1, S60X4, S240X4). Due to the notion prevalent in speech science, that stimulus repetition is detrimental to the transfer of learning (e.g., Greenspan et al., 1988) the remaining planned comparisons were also devoted to the effects of the training set size. Each of those focused on two groups with the same amount of training to account for the possibility that the effect of set size depends on the overall length of training.
4) The effect of set size for brief training sessions was tested by comparing the two groups that received brief training sessions with the two different set sizes (L60X4 vs. S60X4).
5) Likewise, the effect of set size for more intense training sessions was tested by comparing the two groups that received long training sessions with either large (L240X4) or small (S240X4) training sets.

For the ANOVAs, Levene tests were used to determine homogeneity of variance across groups and degrees of freedom were adjusted accordingly in cases of violation. Partial eta-squared (η^2_p) was used to estimate the ANOVA group effect sizes. For the planned comparisons, effect sizes are reported using Cohen's d. Although these comparisons were all pre-planned, not all comparisons were orthogonal. Therefore, to account for multiple comparisons, critical alphas were adjusted with Bonferroni corrections based on the number of times data from each group was used in the planned comparisons (see Table 2). These analyses were carried out in SPSS (v.23).

A Bayes factor corresponding to each of the ANOVAs (assuming equal priors) was calculated in JASP (JASP team, 2018). Bayes factors for the planned comparisons were calculated using the calculator provided by Dienes (2008). Bayes factors are the ratios of the likelihoods of two competing hypotheses (H0, that there is no group difference and H1 that the groups differ) given the data (Dienes, 2008). In the current study, H1 was modelled based on the training effects obtained in our previous studies (relative to untrained controls) (Banai & Lavner, 2014; Gabay et al., 2017; Tarabeih-Ghanayim, Lavner & Banai, 2019). We reasoned that the effect of any of the training related variables cannot be larger than the overall effect of training which we estimated at 25% for learning and acoustic transfer and 15% for semantic transfer. This

Table 2 Planned comparisons

Contrast number	Effect	Groups	Critical alpha[e]
1	Training	Control vs. all trained	0.05
2	Spacing[a]	L240X1 & S240X1 vs. L60X4 & S60X4	0.025
3	Intensity[b]	L60X4 & S60X4 vs. L240X4 & S240X4	0.025
4	Set-size	L240X1, L60X4 & L240X4 vs. S240X1, S60X4, S240X4	0.017
5	Set-size[c] (brief sessions)	L60X4 vs. S60X4	0.017
6	Set-size[d] (intense sessions)	L240X4 vs. S240X4	0.017

[a] All groups received the same total number of training trials, therefore this contrast allows teasing out the effect of distributing a given number of trials across a number of sessions

[b] Groups with different numbers of total training trials are compared to estimate the effect of training intensity, collapsing across set size

[c,d] These test the effect of set size for groups with the same numbers of training trials

[e] Critical alpha adjusted where necessary using Bonferroni correction

calculation is different from the more standard approach in which no explicit assumptions about the size of the expected effect are made, but we found it appropriate for two reasons: First, as explained above, we expected semantic transfer effects to be relatively small. Therefore, it seems reasonable to base the calculation on realistic expectations based on previous data. Second, as replications often result in smaller effect sizes than the original studies (see, e.g., Dienes & Mclatchie, 2018), constraining the calculation based on previous data is still quite conservative because if the group effect in the current data is substantially smaller than expected, the resulting Bayes factors are also expected to be small.

Bayes factors larger than 3 are considered as evidence for H1; Bayes factors smaller than 1/3 are considered as evidence for H0; values between 1/3 and 3 are cases in which the data provides no real evidence as to the relative probability of H1 and H0 (Dienes & Mclatchie, 2018).

Training phase data is hard to compare across groups due to the differences between the training protocols and thus this data is not presented. Inspection of the learning curves from previous studies of adaptive training on time-compressed speech (Banai & Lavner, 2014, 2016) suggest gradual improvements during training both within and across training sessions.

Results

Baseline performance

Baseline performance (see Fig. 1) was similar across the seven groups of participants despite large variability across individual participants. Consistent with the visual inspection of the data, an analysis of variance with group as independent factor failed to reject the null hypothesis that group means are similar across groups ($F(6,62) = .48$, $p = .82$, $\eta^2_p = 0.045$). A Bayesian analysis was used to compare the relative odds of the two competing hypotheses (H0 that all groups perform similarly and H1 that there are group differences in baseline performance) given the observed baseline data. The resulting Bayes Factor was small ($BF_{10} = 0.086$), which is considered strong evidence in favor of the null hypothesis. Therefore, group differences in test-phase performance can be attributed to training rather than to initial differences in the recognition of time-compressed speech.

Test phase performance

Training-induced learning and transfer were estimated by comparing test-phase performance across groups. Group means and confidence intervals are presented in Fig. 2. Descriptively, performance appears poorest with massed training (shown with triangles) and best with intensive distributed training (shown in diamonds) with intermediate performance

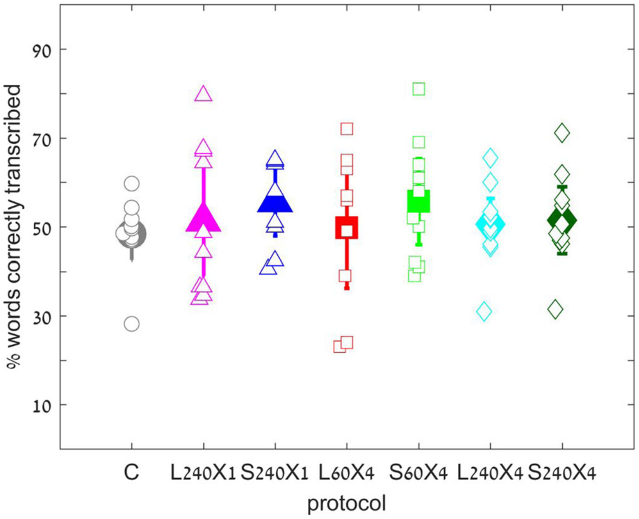

Fig. 1 Baseline performance across groups. Means (filled symbols) and 95% confidence intervals are shown for each group. Individual data are shown in smaller, unfilled symbols

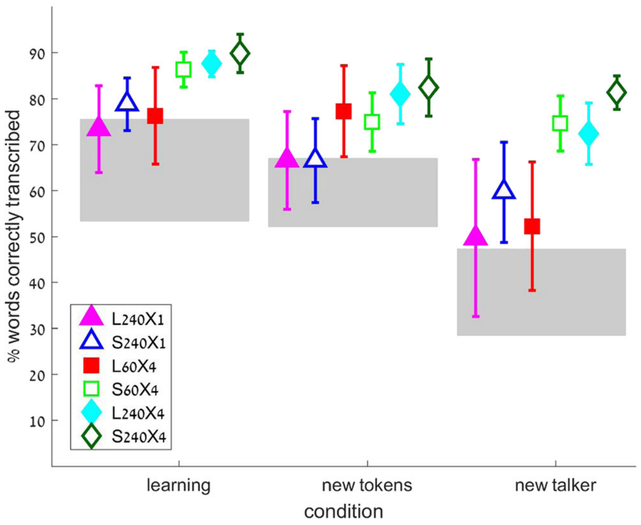

Fig. 2 Test-phase performance by training protocol and condition. **Left to right:** Learning (performance with sentences encountered in baseline and training), transfer to new tokens (performance with new sentences presented by the familiar talker) and transfer to a new talker (performance with sentences encountered in baseline and training presented by the unfamiliar talker). For each condition, the 95% confidence interval of the control group is marked by a gray rectangle. For each trained group, mean and 95% confidence interval are shown. The two single session groups (triangles) are shown to the left of each panel, followed by the two groups who received distributed training with brief sessions (squares) then those who received distributed training with intense sessions (diamonds). Empty and filled symbols mark groups that practiced with the small set size and large set size, respectively

following brief distributed training (squares). This observation is also consistent with Fig. 3, which shows training effect sizes relative to the untrained control group, which also appear to increase from left to right. Statistical analyses were carried out on the data from each test condition, focusing on the planned comparisons between training groups (Table 2).

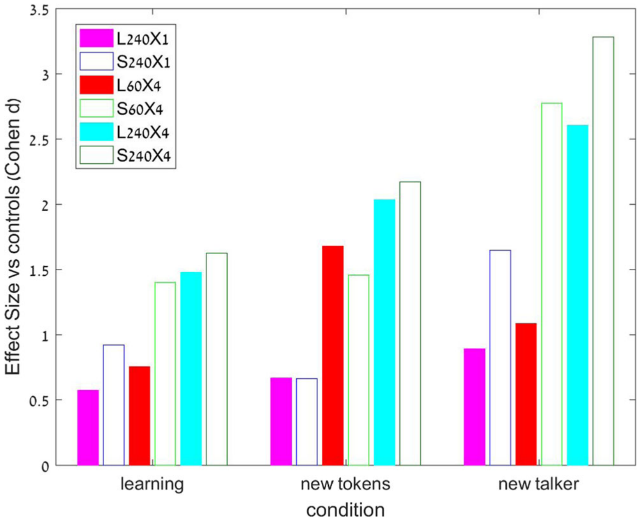

Fig. 3 Learning and transfer effect sizes. For each of the trained groups, Cohen's d relative to untrained controls is shown. For the order of the groups (left to right, see Fig. 2)

Learning Performance on the learning condition is shown on the left side of Fig. 2 and learning effect sizes (relative to untrained controls) are shown on the left side of Fig. 3. Generally, trained groups had higher speech recognition than untrained controls, with medium (0.5) or higher effect sizes. One-way ANOVA with group as independent factor and recognition of the previously encountered sentences as dependent factor yielded a significant group effect (F(6,62) = 8.02, p < .001, η^2_p = 0.437), which was also supported by the Bayesian analysis (BF_{10} = 3.383e+6). Planned comparisons suggested first that the effect of training (all trained groups vs. untrained controls) was significant with a strong effect size (t(62) = 4.88, p < 0.001, Cohen's d = 1.33). The Bayesian analysis also provided strong evidence for the training effect (BF = 53,418). Second, the overall effect of spacing the training experience over multiple sessions was not significant (L240X1 and S240X1 vs. L60X4 and S60X4; t(62) = 1.61, p = .113, Cohen's d = 0.52), and the Bayesian analysis failed to provide conclusive support for H1 over H0 (BF = 1.29). Third, longer training sessions (i.e., the intensity effect) yielded greater learning than shorter ones (L240X4 and S240X4 vs. L60X4 and S60X4; t(62) = 2.47, p = .016, Cohen's d = 0.84, BF = 6.89). Finally, the overall effect of set size was not significant once multiple comparisons accounted for (t(62, = 2.24, p = .028, Cohen's d = 0.54), but the Bayesian analysis provided evidence that practice with a small set size may have resulted in greater learning than practice with a large set size (BF = 3.28). The outcomes of the analyses in which the effect of set-size was assessed for either brief (Fig. 2, left panel, L60X4 vs. S60X4, t(62) = 2.18, p = .033, Cohen's d = 0.99) or intense training sessions (Fig. 2, left, L240X4 vs. S240X4, t(62) = 0.68, p = .501, Cohen's d = 0.45) were consistent with those of the overall analysis. The Bayesian analysis again provided some evidence of a set size effect with brief (BF = 5.2) but not long (BF = 0.45) training sessions (note, however, the high level of performance in the two groups who received distributed training with long training sessions).

Transfer to new tokens (semantic transfer) Performance on the new tokens condition (Figs. 2 and 3, middle section) also differed across groups (F(6,62) = 5.63, p < 0.001, η^2_p = 0.353; BF_{10} = 1,829). As for learning, trained groups recognized new tokens more accurately than the untrained control group with medium to large effect sizes. Furthermore, a visual inspection of Fig. 2 suggests that performance improved as training was distributed across sessions.

Planned comparisons suggested that the effect of training (all trained groups vs. untrained controls) was significant (t(62) = 3.89, p < .001, Cohen's d = 1.33, BF = 585). Consistent with the visual inspection, distributing the training over four sessions yielded more transfer to new tokens than massing it in a single session (L240X1 and S240X1 vs.

L60X4 and S60X4; 6(62) = 2.45, p = .015, Cohen's d = 0.74, BF = 11.77). Also consistent with the visual inspection, intensity and training set-size had no effect on transfer to new tokens (intensity effect: t(62) = 1.65, p = .104, Cohen's d = 0.57, BF = 2.24; overall set-size effect: t(62) = -0.24, p = .812, Cohen's d = 0.04, BF = 0.26; set-size effect with brief training sessions: t(62) = 0.59, p = .558, Cohen's d = 0.26, BF = 0.74; set- size effect with more intense training sessions: t(62) = -0.32, p = .750; Cohen's d = -0.16, BF = 0.55).

Cross-talker transfer (acoustic transfer) As shown in Fig. 2 (right side), performance with the unfamiliar talker appears poorer than in the other conditions, but still different across groups (F(6,62) = 11.79, p < .001, η^2_p = 0.533, BF_{10} = 8.609e+7). Trained groups tended to outperform the untrained group with medium to large effect sizes (see Fig. 3).

Planned comparisons (with degrees of freedom corrected because equal variances could not be assumed across groups) suggested that the effect of training was significant (t(13) = 5.88, p < .001, Cohen's d = 1.70, BF 6.199e+6). Furthermore, although there were no statistical evidence that distributing training across four sessions contributed to performance (t(25) = 1.62, p = .117, Cohen's d = 0.53, BF = 2.04), training intensity (t(19) = 3.67, p = .002, Cohen's d = 0.91, BF = 284) and set size both had an effect. Overall, training with a smaller training set yielded greater transfer to the unfamiliar talker than training with a larger training set (t(30) = 3.59, p = .001, Cohen's d = 0.81, BF = 229). This was true regardless of training intensity (brief sessions: t(11) = -3.40, p = .006, Cohen's d = 1.58, BF = 149; long sessions: t(16) = -2.58, p = .020, Cohen's d = 1.11, BF = 9.68).

Discussion

The present findings are consistent with two general conclusions. First, the perceptual learning of time-compressed speech is robust, because at test, trained listeners outperformed the untrained control group across test conditions and regardless of the training protocol. Second, although the number of different tokens encountered during training and the length of each training session influenced learning and its acoustic transfer (to a new talker repeating familiar sentences), these manipulations had no effect on the critical test of semantic transfer (new tokens produced by the trained talker). Of the three training-related variables we considered, spacing was the only one that had an effect on transfer to new tokens. Distributing training over a number of sessions resulted in greater cross-token transfer than massing the same number of training trials during a single session. In the following paragraphs, these aspects of the findings are discussed and the implications for the utility of training for clinical or educational purposes are considered.

In the current study, when multi-session training was provided, longer training sessions resulted in greater learning of the trained stimuli and in greater cross-talker transfer, but not in greater cross-token transfer. It thus seems that simply providing more training time on each session was not sufficient to drive listeners to rely on low-level speech cues in lieu of their default reliance on high-level cues (Ahissar et al., 2009; Mattys et al., 2009). This finding extends previous observations from auditory discrimination training that beyond a certain threshold, providing additional daily training was not advantageous and may have actually slowed learning (Molloy, Moore, Sohoglu, & Amitay, 2012; Wright & Sabin, 2007). It is also consistent with the notion that perceptual learning is specific to the physical characteristics of the trained stimuli and that further practice with the same stimuli might actually result in greater learning specificity (Greenspan et al., 1988; Hussain et al., 2012). On the other hand, for a total given number of trials, distributed training resulted in greater transfer to new tokens than massed training, and this was true both with and without stimulus repetition during training. This finding suggests that the previous observations that generalization lagged behind learning and required more practice sessions to emerge (Banai & Lavner, 2014; Wright, Wilson, & Sabin, 2010) are less likely attributable to the overall number of training trials and more likely associated with the distribution of practice.

Distributed training, learning, and transfer

Finding that of the different training protocols, distributed training was the only one that contributed to cross-token transfer is meaningful because transfer is considered the hall mark of speech learning, and because the application of speech training for education or rehabilitation depends on transfer. Although this would have been expected given the vast literature on the effects of distributed training on recall in verbal learning studies (Cepeda et al., 2006), the current finding demonstrates that spacing can support the transfer of learning even with no direct effect on learning itself. This is consistent with the suggestion that auditory perceptual learning and transfer operate by partially distinct mechanisms (Wright et al., 2010). Furthermore, to our knowledge this is the first demonstration of the contribution of distributed training to the perceptual learning of speech because a previous study on the perceptual learning of speech in noise (Tye-Murray et al., 2017) found no differences between the distributed and massed protocols. Due to the large number of differences between this study and ours (including the type of perceptually difficult speech used, the study population, the overall length of the training protocol and the tasks performed during training), further research is required to determine the replicability and extent of the distributed training effect reported here.

Why would distributed training improve the transfer of learning to new tokens? According to the Reverse Hierarchy Theory (RHT, see *Introduction*), cross-token transfer requires the use of the low-level cues associated with time compression. These cues are not readily accessible to listeners with no prior experience with compressed speech, but can become usable with certain types of practice (Ahissar et al., 2009; Francis et al., 2007). That distributed training resulted in more transfer than massed training suggests that of the training protocols used here, distributed practice provides the conditions necessary to make low-level cues most accessible at test. Consistently, previous studies showed that multi-session (i.e., distributed) training with whole words (presented as synthetic or noise-vocoded speech) lead to improved identification of individual consonants (Francis et al., 2007; Stacey & Summerfield, 2008). One reason that distributed training was the most effective in doing so might be that it provides learners multiple opportunities to consolidate and reconsolidate learning (Fenn et al., 2003; Karni & Sagi, 1993), such that after training the relevant low-level cues were more accessible than in the other training protocols. Indeed, for accented speech, where cross-talker differences limit the transfer of learning, transfer to a phonetically distinct talker was larger in a group who had the opportunity to consolidate learning overnight than in a group that was re-tested on the same day (Xie, Earle, & Myers, 2018). Similarly, the cross-modal transfer of temporal discrimination gains associated with practice was also larger after overnight consolidation (Bratzke, Schroter, & Ulrich, 2014).

We note that in the current study the training-to-test interval was longer for the groups that completed massed training (6–7 days) than for the groups that completed distributed training (2–3 days). This was because we required that all participants complete the study within a fixed period of time, and also that there be a minimal spacing of the training sessions in the distributed training groups. Therefore, it could be argued that the effect of distributed training stemmed not from the distribution of training, but rather due to greater decay of learning in the massed training groups. We think this is unlikely given that time-compressed speech learning has been shown to maintain for substantially longer periods of time (even up to a year, Altmann & Young, 1993). Comparisons of data from previous studies in which the overall design was similar to that of the current study also suggests similar learning and transfer effect sizes when 1–10 days have elapsed between training and testing (Banai & Lavner, 2012, 2014, 2016; Manheim et al., 2018).

Stimulus repetition during training, learning, and transfer

In the present study, the number of different sentences encountered during training had little effect on the transfer of learning to new tokens. However, and consistent with the notion that small stimulus sets increase the specificity of learning (Greenspan et al., 1988; Hussain et al., 2012), the groups that experienced only 60 different sentences during training exhibited (according to the Bayesian analysis) more learning of the trained items and more transfer when trained items were presented by a new talker than the groups that experienced 240 different sentences. Note, however, that because the talkers in the current study were both men with similar voice pitch, further studies with more acoustically distinct talkers are required to test the generality of this finding. As for semantic transfer, while we hesitate to offer a strong interpretation for a null finding, several differences can account for why here, practice with a larger number of different sentences had no effect on the token specificity of learning.

The main difference between our study and previous studies that served the basis of the notion that item variability positively contributes to generalization (Baese-Berk et al., 2013; Bradlow & Bent, 2008) is that we increased the number different stimuli without increasing the number of different talkers. As explained in the *Introduction*, this was motivated by previous findings that learning of time-compressed speech has a token-specific component even for the trained talker. Data also suggest that for time-compressed speech, transfer to new tokens presented with new acoustics is limited (Manheim et al., 2018; Tarabeih-Ghanayim, Lavner, & Banai, 2019). It had been suggested that wider sampling of the acoustic space of a single talker by increasing the number of sentences for the same talker supports the transfer of learning to new tokens (Greenspan et al., 1988). Nevertheless, this manipulation had no effect on the transfer of learning to new items in our study, suggesting that 60 different sentences may have provided sufficient sampling of the acoustic-phonetic space of the trained talker presented in time-compressed format. To determine whether talker variability diminishes learning specificity (or increases transfer) for new talkers presenting new tokens requires further studies. One previous study (in older adults) suggests that with no stimulus repetition at all and with different talkers encountered during training, resilience to time compression continued to improve throughout 13 training sessions delivered over the course of 4 weeks (Karawani et al., 2015). It could be that the effects of set size thus depend on the overall duration of training. Alternatively, recognition memory due to sentence repetition during practice may have reduced the transfer of learning to new sentences. Note however that for a given training protocol, large set sizes (with no sentence repetition) yielded similar amounts of transfer to new sentences as did small set sizes (with sentence repetition) (see Fig. 3, mid panel). For example, when comparing the two groups who practiced for four brief sessions, the effect sizes of transfer to new tokens were similar whether listeners practiced with 240 different sentences or whether they practiced with 60 different sentences each repeating four times. The same was true for the single session groups. Therefore, while token repetition may result in specific memories for the trained tokens even when presented by a new talker and thus contribute to acoustic transfer, these memories do not seem to interfere with the transfer of learning to new tokens presented by a familiar talker.

Another potential difference between the current and previous studies is that in the past stimulus variability supported learning when the critical parameter could have been easily conceived as categorical. Thus, in studies of non-native phoneme learning, non-native listeners need to learn to classify into different categories based on a feature that is non-contrastive in their first language (e.g., Lively et al., 1993). Similarly, talkers can usually be classified based on their accent or regional dialect (e.g., Clopper & Pisoni, 2004). In these cases, variability may support the transfer of learning by emphasizing the general attributes of the different categories at the expanse of talker- or item- specific attributes (Baese-Berk et al., 2013). However, in the current study, it is hard to conceive of distinct time-compressed categories because during training the level of compression changed adaptively based on performance (see Banai & Amitay, 2015, for further discussion of this issue).

Conclusion

Whereas the very rapid learning of time-compressed speech is highly transferable across items, talkers, compression rates and even languages, the learning that occurs with further practice is more item specific. Of the three training-related parameters considered here, only the distribution of practice across several sessions contributed to the transfer of learning. It remains to be seen whether other manipulations such as increasing the number of different talkers or spacing training over a greater number of sessions influences learning specificity, but as it currently stands the observed specificity limits the utility of time-compressed speech training for real-world applications.

Acknowledgements This study was supported by the National Institute for Psychobiology in Israel – founded by the Charles E. Smith Family (grant 108-14-15). We thank Osnat (Ossi) Teller for help with participant recruitment and data collection.

Open Practices Statement The data on which this article is based will be made available as Supplemental Material. Stimulus materials and code are available upon request. The study was not preregistered.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest

References

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *Journal of the Acoustical Society of America, 126*(5), 2649-2659. doi:https://doi.org/10.1121/1.3216914

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 364*(1515), 285-299. doi: https://doi.org/10.1098/rstb.2008.0253

Altmann, T.M., & Young, D. (1993). *Factors affecting adaptation to time-compressed speech*. Paper presented at the EUROSPEECH '93, Berlin.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America, 133*(3), EL174-180. doi: https://doi.org/10.1121/1.4789864

Banai, K., & Amitay, S. (2015). The effects of stimulus variability on the perceptual learning of speech and non-speech stimuli. *PloS One, 10*(2), e0118465. doi: https://doi.org/10.1371/journal.pone.0118465

Banai, K., & Lavner, Y. (2012). Perceptual learning of time-compressed speech: more than rapid adaptation. *PloS One, 7*(10), e47099. doi: https://doi.org/10.1371/journal.pone.0047099

Banai, K., & Lavner, Y. (2014). The effects of training length on the perceptual learning of time-compressed speech and its generalization. *Journal of the Acoustical Society of America, 136*(4), 1908. doi: https://doi.org/10.1121/1.4895684

Banai, K., & Lavner, Y. (2016). The effects of exposure and training on the perception of time-compressed speech in native versus nonnative listeners. *Journal of the Acoustical Society of America, 140*(3), 1686. doi: https://doi.org/10.1121/1.4962499

Borrie, S. A., Lansford, K. L., & Barrett, T. S. (2017). Generalized Adaptation to Dysarthric Speech. *Journal of Speech, Language, and Hearing Research, 60*(11), 3110-3117. doi: https://doi.org/10.1044/2017_JSLHR-S-17-0127

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729. doi: https://doi.org/10.1016/j.cognition.2007.04.005

Bratzke, D., Schroter, H., & Ulrich, R. (2014). The role of consolidation for perceptual learning in temporal discrimination within and across modalities. *Acta Psychologica, 147*, 75-79. doi: https://doi.org/10.1016/j.actpsy.2013.06.018

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354-380. doi: https://doi.org/10.1037/0033-2909.132.3.354

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47*(Pt 3), 207-239.

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), 222-241. doi: https://doi.org/10.1037/0096-3445.134.2.222

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan. Website for associated online Bayes factor calculator: http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm

Dienses, Z., & Mclatchie, N. (2018). Four reasons to prefer a Bayesian analysis over significance testing. *Psychonomic Bulleting & Review, 25(1)*:207-218.

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), 914-927.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425*(6958), 614-616. doi: https://doi.org/10.1038/nature01951

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287*(5777), 43-44.

Francis, A. L., Nusbaum, H. C., & Fenn, K. (2007). Effects of training on the acoustic phonetic representation of synthetic speech. *Journal of*

Speech, Language, and Hearing Research, 50(6), 1445-1465. doi: https://doi.org/10.1044/1092-4388(2007/100)

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. doi: https://doi.org/10.1371/journal.pone.0176488

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 421-433.

Henshaw, H., & 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*(5), e62836. doi: https://doi.org/10.1371/journal.pone.0062836

Hussain, Z., McGraw, P. V., Sekuler, A. B., & Bennett, P. J. (2012). The rapid emergence of stimulus specific perceptual learning. *Frontiers in Psychology, 3*, 226. doi: https://doi.org/10.3389/fpsyg.2012.00226

Huyck, J. J., Smith, R. H., Hawkins, S., & Johnsrude, I. S. (2017). Generalization of Perceptual Learning of Degraded Speech Across Talkers. *Journal of Speech, Language, and Hearing Research, 60*(11), 3334-3341. doi: https://doi.org/10.1044/2017_JSLHR-H-16-0300

JASP team (2018). JASP(Version 0.8.6). [Computer software].

Karawani, H., Bitan, T., Attias, J., & Banai, K. (2015). Auditory Perceptual Learning in Adults with and without Age-Related Hearing Loss. *Frontiers in Psychology, 6*, 2066. doi: https://doi.org/10.3389/fpsyg.2015.02066

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365*(6443), 250-252.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2), 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. *Journal of the Acoustical Society of America, 94*(3 Pt 1), 1242-1255.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2009). Transfer of auditory perceptual learning with spectrally reduced speech to speech and nonspeech tasks: implications for cochlear implants. *Ear and Hearing, 30*(6), 662-674. doi: https://doi.org/10.1097/AUD.0b013e3181b9c92d

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

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203-243. doi: https://doi.org/10.1016/j.cogpsych.2009.04.001

Molloy, K., Moore, D. R., Sohoglu, E., & Amitay, S. (2012). Less is more: latent learning is maximized by shorter training sessions in auditory perceptual learning. *PloS One, 7*(5), e36929. doi: https://doi.org/10.1371/journal.pone.0036929

Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS Biology, 6*(5), e126. doi:https://doi.org/10.1371/journal.pbio.0060126

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. doi: https://doi.org/10.1016/j.visres.2009.09.004

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1315-1330. doi: https://doi.org/10.1037/0096-1523.31.6.1315

Prior, A., & Bentin, S. (2006). Differential integration efforts of mandatory and optional sentence constituents. *Psychophysiology, 43*(5), 440-449. doi: https://doi.org/10.1111/j.1469-8986.2006.00426.x

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Atten Percept Psychophys, 71*(6), 1207-1218. doi: https://doi.org/10.3758/APP.71.6.1207

Saunders, G. H., Smith, S. L., Chisolm, T. H., Frederick, M. T., McArdle, R. A., & Wilson, R. H. (2016). A Randomized Control Trial: Supplementing Hearing Aid Use with Listening and Communication Enhancement (LACE) Auditory Training. *Ear and Hearing, 37*(4), 381-396. doi: https://doi.org/10.1097/AUD.0000000000000283

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: biological mechanisms. *Cerebral Cortex, 22*(5), 1180-1190. doi: https://doi.org/10.1093/cercor/bhr196

Stacey, P. C., & Summerfield, A. Q. (2008). Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *Journal of Speech Language and Hearing Research 51*(2), 526-538. doi: https://doi.org/10.1044/1092-4388(2008/038)

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28*(3), 455-462.

Svirsky, M. A., Talavage, T. M., Sinha, S., Neuburger, H., & Azadpour, M. (2015). Gradual adaptation to auditory frequency mismatch. *Hearing Research, 322*, 163-170. doi: https://doi.org/10.1016/j.heares.2014.10.008

Tarabeih-Ghanayim, M., Lavner, Y., & Banai, K. (2019). *Tasks, talkers and the perceptual learning of time-compressed speech*. Poster presented at the 42nd midwinter meeting of the Association for Research in Otolaryngology, Baltimore, MD.

Tye-Murray, N., Spehar, B., Barcroft, J., & Sommers, M. (2017). Auditory Training for Adults Who Have Hearing Loss: A Comparison of Spaced Versus Massed Practice Schedules. *Journal of Speech, Language, and Hearing Research, 60*(8), 2337-2345. doi: https://doi.org/10.1044/2017_JSLHR-H-16-0154

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

Wright, B. A., & Sabin, A. T. (2007). Perceptual learning: how much daily training is enough?. *Experimental Brain Research.*

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. doi: https://doi.org/10.1523/JNEUROSCI.1441-10.2010

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.