

Rapid Perceptual Learning: A Potential Source of Individual Differences in Speech Perception Under Adverse Conditions?

Trends in Hearing
Volume 24: 1–16
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216520930541
journals.sagepub.com/home/tia



Tali Rotman¹, Limor Lavie¹, and Karen Banai¹ 

Abstract

Challenging listening situations (e.g., when speech is rapid or noisy) result in substantial individual differences in speech perception. We propose that rapid auditory perceptual learning is one of the factors contributing to those individual differences. To explore this proposal, we assessed rapid perceptual learning of time-compressed speech in young adults with normal hearing and in older adults with age-related hearing loss. We also assessed the contribution of this learning as well as that of hearing and cognition (vocabulary, working memory, and selective attention) to the recognition of natural-fast speech (NFS; both groups) and speech in noise (younger adults). In young adults, rapid learning and vocabulary were significant predictors of NFS and speech in noise recognition. In older adults, hearing thresholds, vocabulary, and rapid learning were significant predictors of NFS recognition. In both groups, models that included learning fitted the speech data better than models that did not include learning. Therefore, under adverse conditions, rapid learning may be one of the skills listeners could employ to support speech recognition.

Keywords

speech recognition, auditory learning, age-related hearing loss, fast speech

Received 14 February 2020; Revised 20 April 2020; accepted 6 May 2020

Dynamic matching of incoming speech input with pre-existing phonological and lexical representations facilitates speech perception, especially under suboptimal or adverse conditions (Davis & Johnsruide, 2007; Guediche et al., 2014; Mattys et al., 2009; Samuel, 2011). However, such suboptimal conditions also yield substantial individual differences in speech recognition which are only partially explained by various sensory (e.g., hearing acuity), cognitive (e.g., working memory and vocabulary), and demographic factors (Benichov et al., 2012; Bent et al., 2016; Carbonell, 2017; DeCaro et al., 2016; Gordon-Salant & Fitzgibbons, 1997; Humes & Dubno, 2010; Mattys et al., 2012; McLaughlin et al., 2018; Nagaraj, 2017; Wingfield & Tun, 2001). Another potential source of individual differences in speech recognition which we have recently revealed is perceptual learning (Banai & Lavie, 2020; Karawani et al., 2017; Manheim et al., 2018), defined as experience-induced changes in the perception of stimulus arrays (Green et al., 2018; Samuel & Kraljic, 2009). These studies showed high correlations between rapid perceptual learning of

time-compressed speech (TCS) and the recognition of natural-fast speech (NFS) in both younger and older adults. However, these correlations could reflect the contribution of the sensory and cognitive factors already known to associate with speech recognition, and not a unique contribution of learning. Therefore, the goal of the current study was to test the hypothesis that in normal-hearing young adults as well as in older adults with hearing loss, rapid auditory perceptual learning is a unique predictor of speech recognition under adverse conditions while accounting for other sensory and cognitive variables. To this end, we assessed rapid learning of TCS and the recognition of NFS in younger and older adults, speech perception in noise (young adults only), and indices

¹Department of Communication Sciences and Disorders, University of Haifa

Corresponding Author:

Karen Banai, Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel.
Email: kbanai@research.haifa.ac.il



of cognitive performance (selective attention, working memory span, vocabulary, and nonverbal intelligence).

Perceptual Learning and the Recognition of Degraded Speech

Samuel and Kraljic (2009) suggested that a primary purpose of perceptual learning for speech in adulthood is to allow listeners to understand degraded speech (speech that deviates from the norm or that is presented under adverse conditions). Consistent with this suggestion, multiple labs reported improved recognition of degraded speech (e.g., noisy, vocoded, time-compressed, accented) following either brief experiences or longer training (e.g., Adank & Janse, 2009; Baese-Berk et al., 2013; Banai & Lavner, 2014; Bradlow & Bent, 2008; Clopper & Pisoni, 2004; Davis et al., 2005; Dupoux & Green, 1997; Golomb et al., 2007; Greenspan et al., 1988; Stacey & Summerfield, 2008). The current study focuses on rapid learning that follows brief encounters with distorted speech. For example, upon first encounter with TCS, substantial improvements occur within minutes of exposure (Altmann & Young, 1993; Dupoux & Green, 1997; Peelle & Wingfield, 2005). The recognition of other forms of degraded speech also improves after brief exposure (Clarke & Garrett, 2004; Davis et al., 2005). However, in studies of perceptual learning, both learning and speech recognition were usually assessed using a single type of speech stimulus (i.e., stimuli were all time-compressed or vocoded). Therefore, the design of those studies makes it hard to determine whether perceptual learning can be viewed as a more general capacity. In our previous studies (Karawani et al., 2017; Manheim et al., 2018), we showed that individual differences in perceptual learning are associated with individual differences in speech recognition assessed with speech materials that are different from those used to assess learning. The rapid occurrence of perceptual learning at least for some forms of degraded speech, (e.g., time-compressed and accented speech, Adank & Janse, 2009; Clarke & Garrett, 2004; Dupoux & Green, 1997; Golomb et al., 2007; Gordon-Salant et al., 2010; Peelle & Wingfield, 2005) together with findings that speech perception is inherently dynamic and flexible (Ahissar et al., 2009; Davis & Johnsrude, 2007; Guediche et al., 2014) lead us to propose that perceptual learning is one of the capacities that serve online speech recognition. If this is the case, the specificity of learning to the type of speech to which listeners are exposed is not an obstacle for learning (in the absence of generalization), but rather, this specificity serves the purpose of coping with the exact characteristics of the encountered degraded speech.

If, as we argue, perceptual learning is a capacity that serves ongoing speech recognition, rapid learning should

explain unique variance in the recognition of different types of distorted speech in addition to the known contributions of other cognitive capacities such as attention, working memory, and vocabulary. When listening to speech, listeners automatically attempt to extract the general meaning of the input (e.g., the identity of a word) and are not attuned to the lower level acoustic and phonetic features of the input (Ahissar et al., 2009; Nahum et al., 2008). However, under adverse conditions, this automatic process often fails, and the failure triggers adaptive processes (Guediche et al., 2014; Sohoglu & Davis, 2016) that allows access to the relevant low-level information. Learning rates and amounts differ across individuals (Manheim et al., 2018; Theodore et al., 2019). We thus argue that because in “good learners” learning is generally quicker and more efficient, they enjoy a general advantage, leading to better perception of degraded speech.

We use the perceptual learning of TCS as an index of rapid learning because most listeners have no experience with this type of speech, making it useful in studying the correlations between learning-as-a-capacity and the recognition of other, more familiar forms of degraded speech (e.g., NFS and speech in noise [SIN]). Although highly compressed speech is initially hard to recognize, recognition accuracy improves quite rapidly with exposure to as few as 10 to 20 time-compressed sentences (e.g., Dupoux & Green, 1997; Golomb et al., 2007). This rapid learning has been shown in young and in older adults as well as in older adults with hearing impairments (Golomb et al., 2007; Manheim et al., 2018; Peelle & Wingfield, 2005). Finally, across age and hearing levels, we previously reported that variance in rapid learning of TCS accounted for unique variance in the recognition of NFS, even after controlling for the known association between the recognition of NFS and TCS (Manheim et al., 2018).

Age-Related Hearing Loss, Perceptual Learning, and the Recognition of Distorted Speech

The recognition of different forms of degraded speech deteriorates with age and age-related hearing loss (Dubno et al., 1984; Gordon-Salant & Fitzgibbons, 1995, 2001; Humes & Christopherson, 1991; Humes et al., 2012; Janse, 2009; Sommers et al., 2020). As with general individual differences in speech recognition, declines in sensory and cognitive factors are not sufficient to explain why some older adults find degraded speech so challenging (Anderson et al., 2013; Peelle & Wingfield, 2016; Wingfield et al., 2005). Rather, it seems that the relative contribution of nonsensory factors may increase with age/hearing loss, perhaps because listeners recruit domain general cognitive resources to compensate for sensory declines (Alain et al., 2004; Bidelman

et al., 2019; Peelle & Wingfield, 2016). Therefore, we studied the contribution of perceptual learning to speech recognition separately in each age-group.

As for learning, perceptual learning for speech is certainly present in older adults with either normal or impaired hearing (Colby et al., 2018; Golomb et al., 2007; Karawani et al., 2016; Manheim et al., 2018; Neger et al., 2014; Peelle & Wingfield, 2005; Schlueter et al., 2016). However, whereas some reported little or no effects of age on learning (Golomb et al., 2007; Peelle & Wingfield, 2005), others found that learning might slow (Manheim et al., 2018) or diminish (Neger et al., 2014) with age. One possibility is that learning is maintained to a greater extent in situations in which lexical information can be used (Colby et al., 2018; Scharenborg & Janse, 2013). Regardless, older adults who maintain better perceptual learning also tend to maintain better speech perception (Karawani et al., 2017; Manheim et al., 2018). Therefore, we now hypothesize that to the extent that perceptual learning constrains speech perception independent of other factors, this should be true regardless of age/hearing.

Cognition and the Recognition of Distorted Speech

Cognitive and linguistic resources are used to support speech recognition under adverse conditions (Gordon-Salant & Fitzgibbons, 1997; McLaughlin et al., 2018; Pichora-Fuller et al., 1995; Wingfield & Tun, 2001). Therefore, in the current study, we assessed vocabulary, working memory, and inhibitory attention, in addition to speech recognition and rapid auditory learning. While an in-depth review of how cognitive and linguistic factors might contribute to speech perception is beyond the scope of the current article, the general notion is that in favorable listening conditions speech perception is an implicit process, through which the acoustic signal is matched with stored representations in long-term memory. On the other hand, signal degradation results in a mismatch that prevents this automatic process, requiring explicit top-down processing (e.g., the Ease of Language Understanding model, Ronnberg et al., 2013, 2019). Consistent with this notion, there are significant associations between aspects of speech processing and working memory (McLaughlin et al., 2018; Pichora-Fuller et al., 1995; Souza & Arehart, 2015). Attention, and particularly the ability to ignore irrelevant information, is also associated with speech perception (Adank & Janse, 2010; Oberfeld & Klockner-Nowotny, 2016; Tierney et al., 2019). Finally, vocabulary has been repeatedly associated with the recognition of different forms of degraded speech (Banks et al., 2015; Bent et al., 2016; McLaughlin et al., 2018), possibly because it reflects

the wealth of previous experiences a listener can draw from to support ongoing speech recognition.

The Current Study

The major goal of this study was to test facts of the hypothesis that perceptual learning (as measured with a TCS task) explains individual differences in independent estimates of speech recognition when hearing level and cognitive factors are considered. In young adults, we estimated, in addition to rapid learning, the recognition of NFS and SIN. In older adults with age-related hearing loss, we repeated the same assessment with two major differences: First, only NFS was tested. Second, speech materials were amplified, and the rate of TCS was slower to obtain an estimate of rapid learning that is as free as possible from sensory effects.

Methods

Older adults with age-related hearing loss and young adults with normal hearing participated in this study. They were tested on indices of hearing, rapid perceptual learning, speech recognition, and cognition. However, given the known effects of age/hearing loss on some of these processes, test protocols differed in the two groups: First, we adapted the TCS task used to assess rapid learning such that slower speech rates were presented to older adults. Older adults were also given more trials on this task. This was done so that the assessments of the relationships between learning on one hand and speech recognition on the other are as clean as possible from the deleterious effects of age/hearing loss on rapid learning of TCS with this learning paradigm (e.g., Manheim et al., 2018). Second, young adults were tested on three conditions of challenging speech: two conditions of NFS and one condition of SIN. However, due to time constraints (both in testing individual participants and in the time devoted to pilot testing to establish appropriate signal-to-noise ratios [SNRs] for our stimuli in noise), in older adults, speech recognition was assessed with a single condition of NFS.

Participants

A total of 101 participants, 55 young adults ages 18 to 34 years (mean age: 24 ± 4) and 46 older adults ages 65 to 89 years (mean age: 75 ± 7), were recruited. Young adults were recruited through advertisements at academic institutions and social media; older adults were recruited in audiology clinics, at hospitals and in the community. Inclusion criteria for both younger and older adults were no prior experience with TCS and psychoacoustic testing, high proficiency in Hebrew, normal neurological status (based on self-report), and normal cognitive status (see later). Young adults were paid

participants; older adults received either a pack of hearing aid batteries or a personal TV headset as “thank you” gift.

All young adults (20 males, 35 females) had normal hearing (mean 4 frequency [0.5, 1, 2, and 4 kHz] pure-tone average [PTA] in both ears 5 ± 3 dB HL, range 1–12 dB), and all had high school education or higher (see later). The older adults were all with age-related hearing loss (mean 4 frequency PTA in both ears 50 ± 8 dB HL, range 30–70 dB; see Figure 1 for audiograms). Inclusion criterion for suprathreshold recognition score was $\geq 60\%$, actual range was 72% to 100%, and mean score was $90\% \pm 7\%$. Cognitive status was measured with the Hebrew version of the Mini-Mental State Examination (MMSE; Folstein et al., 1975), inclusion criterion: ≥ 24 , mean score 28 ± 2 . All but 4 participants had high school education or higher, range 8 to 22 years, mean 14 ± 3 . One participant was excluded from the study for not following the inclusion criterion for MMSE (score ≥ 24).

Twenty-three of the participants were experienced with hearing aids for at least 1 year, and 22 had similar hearing but no experience with amplification devices. Both groups were counterbalanced in mean age, years of education, and MMSE score. Preliminary analysis suggested that (unaided) hearing, speech perception, learning, and cognition did not differ between the two groups. Therefore, for the purpose of the current report, data were collapsed across the two groups. Note that these data have been presented at the 2019 International Symposium on Auditory and Audiological Research, and a brief report has been published in the symposium’s proceedings (Rotman et al., 2020).

Participants were compensated for their time. 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 (approval number 362/18).

Stimuli

Sixty-two simple Hebrew sentences, 5- to 6-word long (adapted from Prior & Bentin, 2006) with a common subject–verb–object grammatical structure were taken from a set of prerecorded Hebrew sentences. Half of the sentences were semantically plausible (e.g., “The young woman braids her long hair”), and the other half were semantically implausible (e.g., “The broken window plays an electric guitar,” which is implausible because windows do not play guitars). Note that plausibility was based on the semantic congruency of the content and not on predictability or grammatical agreement.

The presentation order of plausible and implausible sentences was random throughout the different tasks. The sentences used for testing speech perception resembled those used for learning evaluation in their length, grammatical structure, semantic plausibility, and recording method and conditions. All sentences were recorded in a sound attenuating booth and sampled at 44 kHz using a built-in MacBook Air microphone and converted to WAV format by three talkers (all female native speakers of Hebrew). The root-mean-square levels of the sentences were normalized after recording using Audacity audio software.

Speech Perception and Rapid Learning Tasks

Speech perception was evaluated with NFS (both groups) and with speech in 4-talker bubble noise (young adults only). NFS was recorded by two different female native speakers of Hebrew. Talker 1 recorded 20 different sentences at an average rate of 221 words/min ($SD = 23$); Talker 2 recorded 10 sentences at a rate of 189 words/min ($SD = 30$). In addition, 10 sentences were recorded by Talker 3 at 113 ± 12 words/min and mixed with 4-talker bubble noise ($SNR = -5$ dB). This SNR was determined based on a previous study in which the same recordings were used on a different cohort of young adults (Banai & Lavie, 2020). Young adults were tested with NFS of the two talkers (1 and 2), 10 sentences each, and with 10 SIN sentences. Older adults were tested with 20 natural-fast sentences of Talker 1 (221 words/min). Talker 2 was added for young adults because pilot testing suggested that Talker 1 (despite being faster) was hardly challenging for young adults. We note that Talker 1 is considered more pleasant and clear than talker two, perhaps because she is a clinical audiologist and is used to speaking clearly.

Rapid learning was evaluated with a TCS task. Young adults listened to 10 sentences; older adults to

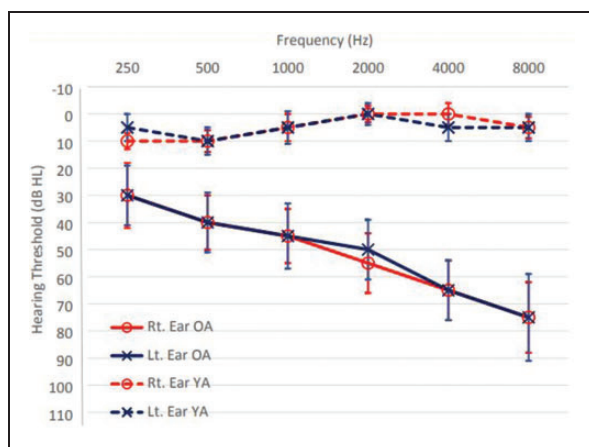


Figure 1. Mean Audiograms. Mean thresholds and standard deviations are shown; Older adults (OA) audiogram in full lines, young adults (YA) in dashed lines.

20 sentences. Sentences (recorded by Talker 3) were compressed to 30% of their duration for the young adults and to 50% of their duration for the older group. Time compression was applied in MATLAB, using a Wave form Similarity Overlap and Add algorithm that modifies the rate while preserving other qualities of the speech stimuli, such as pitch and timbre (Verhelst & Roelands, 1993). Because recognition of TCS declines with age and hearing loss (e.g., Gordon-Salant & Fitzgibbons, 2001; Letowski & Poch, 1996; Wingfield et al., 1985), a fair assessment of learning requires giving older adults speech at a rate that is not as hard as to prohibit learning. Therefore, based on our previous studies with similar sentences (e.g., Banai & Lavner, 2014; Manheim et al., 2018), speech was compressed to 30% of its natural duration for young adults and to 50% of its natural duration for older adults. In Manheim et al., we used a compression rate of 40% for old hearing-impaired adults. However, because in the current study the hearing losses of the participants were more severe, and the sample was somewhat older (on average), a small-scale pilot study (8 participants) was conducted to set the compression rate for the older hearing-impaired adults. The 40% rate yielded a floor effect, and the 50% criterion was deemed adequate. In addition, the number of words per minute in the 50% compression rate matched the mean rate of the NFS presented to the older group (221 words/min). Note however that this adjustment resulted in substantial differences in baseline performance between younger and older adults (see Table 1), which is one of the reasons data was not directly compared across groups.

Sentences were presented (to both ears) through Sennheiser HD-215 headphones (hearing-aids were not used during stimuli presentation), at the most comfortable level (MCL) of each listener. The MCLs were

determined in reference to the levels in dB SPL for a 1-kHz pure tone. The headphones' output was matched to the individual MCL with a sound level meter.

After hearing each sentence once, the younger participants were asked to write it down as accurately as possible. Participants in the older group were asked to repeat what they heard after each single trial, and the experimenter transcribed their replies. Each sentence could be played only once, and no feedback was provided. Performance was scored off-line. All words, including function words, were counted for scoring. Homophonic spelling errors were accepted as correct, consistent with our previous studies (Banai & Lavner, 2014, 2016; Manheim et al., 2018). Other than this exception, words had to be perfectly reported to count as correct.

Cognitive Measures

A computerized version of the flanker task (Eriksen & Eriksen, 1974) was used as a measure of *selective attention*. A computerized version was created in SuperLab, according to the parameters presented by Scharenborg et al. (2015). Participants were seated in front of a white computer screen, with headphones on, and were instructed to indicate (by clicking either the “z” or the “/” key on the keyboard) which direction a middle symbol (an arrow pointing left or right) in a row of five symbols points. There were three types of stimuli: congruent (≫≫≫ or ≪≪≪), incongruent (≫<≫ or ≪>≪), and neutral (==<== or ==>==). Each type (six different sequences) was presented 12 times (a total of 72 trials) in a random order. Each trial started with a beep (a 400 Hz pure-tone signal) and a fixation cross that remained on the screen for 250 ms. Following the fixation cross, the stimulus was presented for 1,500 ms. Participants were instructed to respond in each trial as fast and as accurate as possible. The inter-trial time was 1,000 ms. Before beginning the test, six practice trials were presented. If needed, another six practice trials were given.

The “flanker cost” for each participant was used for statistical analysis. The cost was calculated as the mean logRT (RT = reaction time in ms) of the correct responses in the incongruent trials divided by the mean logRT of the correct responses in the neutral trials. A higher flanker cost (>1) means poorer selective attention. Response accuracies were high (young adults: congruent trials: $M = 98\%$, $SD = 8$; neutral trials: $M = 96$, $SD = 9$; incongruent trials: $M = 92$, $SD = 11$; older adults: congruent: $M = 98$, $SD = 8$; neutral: $M = 97$, $SD = 9$; incongruent: $M = 87$, $SD = 24$). Mean response times in young adults were 1,026 msec ($SD = 96$) in congruent trials, 1,033 msec ($SD = 91$) in neutral trials, and 1,118 msec ($SD = 98$) in incongruent

Table 1. Speech Recognition (Proportions Correct) in Older and Younger Adults.

	Older adults	Younger adults
NFS—Talker 1		
<i>M</i> (<i>SD</i>) [95% CI]	0.36 (0.18) [0.31, 0.41]	0.93 (0.05) [0.92, 0.94]
<i>Mdn</i> (IQR)	0.34 (0.21–0.48)	0.93 (0.91–0.97)
NFS—Talker 2		
<i>M</i> (<i>SD</i>) [95% CI]	–	0.84 (0.07) [0.82, 0.86]
<i>Mdn</i> (IQR)	–	0.85 (0.80–0.90)
SIN		
<i>M</i> (<i>SD</i>) [95% CI]	–	0.62 (0.14) [0.59, 0.66]
<i>Mdn</i> (IQR)	–	0.67 (0.54–0.72)
TCS baseline		
<i>M</i> (<i>SD</i>) [95% CI]	0.37 (0.22) [0.30, 0.44]	0.08 (0.09) [0.06, 0.11]
<i>Mdn</i> (IQR)	0.36 (0.18–0.52)	0.08 (0–0.1)

Note. NFS = natural-fast speech; CI = confidence interval; IQR = interquartile range; SIN = speech in noise; TCS = time-compressed speech.

trials. In older adults, mean response times were 1,199 msec ($SD=112$) in congruent trials, 1,199 msec ($SD=105$) in neutral trials and 1,296 msec ($SD=243$) in incongruent trials.

Three subtests from the Wechsler Adult Intelligence Scale-III in Hebrew were administered: *Working memory* was tested with the *Digit span* forward and backward subtest; *Vocabulary* test was used to evaluate participants' *semantic knowledge* and verbal concept formation in Hebrew; and *Block design* was used to evaluate *non-verbal reasoning*. Administration and scoring followed the test manual; scaled scores are reported later.

Study Administration and Schedule

Young adults were tested at the Auditory Cognition Lab at the University of Haifa or at their home, according to their preference. Older adults were tested in a quiet room at an audiology clinic or in the participants' home, according to preference. After a short explanation regarding the procedure of the experiment, participants were requested to sign a written informed consent and fill in a short background questionnaire.

Older adults participated in one session (with frequent breaks, as needed) in which MMSE and hearing were tested followed by the speech and cognitive tasks. Younger adults participated in two sessions: In one session, the speech and learning assessments were completed followed by the cognitive tasks. Hearing was tested on another session in which participants also transcribed two time-compressed sentences presented by Talker 3 that was used as baseline for statistical analysis. Although ideally baseline recognition of TCS should have been estimated prior to the estimation of learning, here this assessment in young adults was conducted as part of the hearing assessment that was conducted either before or after the learning assessment. We nevertheless believe that this choice had minimal influence on the findings because even with substantially longer learning experience improvement in the recognition of new sentences produced by a different talker is minimal (e.g., Manheim et al., 2018).

Preliminary Data Analysis

Missing Data and Outliers. One older adult failed to meet the inclusion criteria (scored lower than 24 on the MMSE) and was excluded from data analysis. Therefore, analyses of this group are based on the 45 remaining participants. Five younger adults had missing scores on one of the cognitive tests; therefore, sample size for the prediction models of young adults was 50. Otherwise, all data (including outlying data points) were included in data analysis.

Speech Perception. The number of correct words per sentence per condition was counted for each participant, and the proportion of correctly recognized words out of the total was used for data presentation and statistical modeling. Baseline recognition of TCS was defined as performance with two sentences recorded by Talker 3. These sentences were measured as part of the hearing screening in young adults. For older adults, the first two time-compressed sentences were used.

The semantic plausibility of the sentences could have influenced performance, especially in older adults (e.g., Sheldon et al., 2008). Therefore, the recognition of plausible and implausible sentences was first evaluated separately in older adults. Recognition of the two types of sentences did not differ significantly for either NFS (plausible sentences: $M=0.36$, $SD=0.18$, $Mdn=0.35$, interquartile range [IQR]: 0.20–0.48; implausible sentences: $M=0.36$, $SD=0.18$, $Mdn=0.31$, IQR: 0.24–0.45, Wilcoxon signed-rank test: $W=495$, $p=.60$, Bayesian comparison: $BF_{10}=0.13$) or TCS (plausible sentences: $M=0.45$, $SD=0.25$, $Mdn=0.47$, IQR: 0.25–0.64; implausible sentences: $M=0.48$, $SD=0.24$, $Mdn=0.44$, IQR: 0.28–0.67, Wilcoxon signed-rank test: $W=349$, $p=.97$, Bayesian comparison: $BF_{10}=0.055$). Consequently, analyses in the results section were conducted with no further consideration of semantic plausibility. Younger adults are supposed to be less influenced by semantic plausibility (Sheldon et al., 2008). In addition, in the current study, young adults were tested with fewer sentences than older adults, so we did not separate plausible and implausible sentences.

Rapid Learning. Learning was defined as the rate of improvement in the recognition of TCS over time. It was quantified as the linear slopes of the learning curves over all TCS sentences. Calculation of the slope was based on the linear fit between performance—the proportion of words correctly recognized—in miniblocks of two sentences (approximately 10 words) and the number of the miniblock. Thus, a slope of 0.1 suggests that recognition improved by 10% per miniblock (approximately 1 word per miniblock or 0.5 words/sentence). The decision to use linear slope was based on our previous studies of TCS learning (e.g., Banai & Lavner, 2014). Because we aimed to dissociate between recognition and rapid learning and had no separate baseline estimate for older adults, we used their first two sentences as baseline for data modeling and therefore, for this part of the analysis, slopes were based on 18 sentences only.

Data Availability. Raw data and audio demos can be requested by emailing the corresponding author.

Results

Speech Recognition and Cognition in Younger and Older Participants

Initial exploration of speech perception in the two experiments was based on averaged performance across all sentences at a given condition per participant. As shown in Table 1, older adults found the recognition of naturally fast speech presented by Talker 1 quite challenging. Although Talker 1 was hardly challenging for younger adults, their performance was challenged to a greater extent with the other talker as well as in the SIN task as shown in the right column of Table 1.

Table 2 shows descriptive statistics for the cognitive indices used in the study as well as PTA hearing thresholds that were used as an index of hearing in subsequent analyses. For the tasks taken from the Wechsler Adult Intelligence, scaled scores in both groups fell well within the normal range as described in the test manual.

As expected from the literature, speech perception was correlated with some cognitive indices (see Table 3 for Pearson correlations). On the other hand, perceptual learning slopes were not correlated with any of the cognitive measures (highest correlation was $r = .2$ which was insignificant). Nor were they associated with hearing thresholds ($r = -.2$ in young adults and $r = .002$ in older adults, both insignificant). The correlations between baseline recognition of TCS and the dependent variables led us to include this baseline in the statistical models described later.

Rapid Perceptual Learning of TCS

To determine whether rapid learning has occurred over the course of a brief encounter with TCS, we tracked

recognition accuracy over the course of 10 sentences for young adults and 20 sentences for older adults. Subsequently, the slopes of the learning curves were calculated as explained in the methods. Figure 2 (left column) shows average accuracies of the first five sentences and final five sentences encountered by each group. In young adults, mean recognition accuracies improved from 0.09 in the first 5 sentences to 0.27 in sentences 6 to 10. The group learning slope ($M = .065$, $SD = 0.06$, 95% confidence interval [CI] [0.049, 0.080], Figure 2 midcolumn) was significantly positive with a large effect size ($t(54) = 8.3$, $p < .001$, Cohen's $d = 1.12$ with a 95% CI of 0.78 to 1.45). In older adults, accuracies improved from 0.35 in the first 5 sentences to 0.45 in the next group of 5 and to 0.56 in sentences 16 to 20. Given that compressed speech rates were slower in older adults, it is not surprising that recognition accuracies were quite good. Yet, initial recognition was still low enough to afford improvements. With 20 sentences, learning was significant with a large effect size ($M = .026$, $SD = 0.021$, 95% CI [0.019, 0.032], $t(44) = 8.11$, $p < .001$, Cohen's $d = 1.21$ with a 95% CI of 0.82 to 1.59). These data are consistent with our previous findings in individuals with milder hearing loss (Manheim et al., 2018) and suggest that although learning is present in older adults, it is nevertheless slower than in younger adults with normal hearing. Therefore, in daily situations where older and younger adults are facing similar conditions, perceptual adjustment is expected to be reduced in older adults.

Speech Recognition Versus Rapid Perceptual Learning

As shown in Figure 3, speech recognition was associated with perceptual learning (assessed as slope over 10

Table 2. Estimates of Hearing and Cognition.

	Older adults	Younger adults
Hearing (PTA4, dB)		
<i>M</i> (<i>SD</i>)	50 (8)	5 (3)
<i>Mdn</i> (<i>IQR</i>)	49 (44–56)	5 (3–7)
Vocabulary (scaled score)		
<i>M</i> (<i>SD</i>)	9 (2)	12 (2)
<i>Mdn</i> (<i>IQR</i>)	9 (7–11)	12 (11–13)
Digit span (scaled score)		
<i>M</i> (<i>SD</i>)	9 (2)	10 (2)
<i>Mdn</i> (<i>IQR</i>)	9 (7–11)	10 (8–12)
Block design (scaled score)		
<i>M</i> (<i>SD</i>)	9 (2)	12 (2)
<i>Mdn</i> (<i>IQR</i>)	9 (7–12)	12 (10–14)
Flanker cost		
<i>M</i> (<i>SD</i>)	0.99 (0.15)	1.00 (0.03)
<i>Mdn</i> (<i>IQR</i>)	1.01 (1.008–1.016)	1.01 (1.002–1.014)

Note. PTA = pure-tone average; IQR = interquartile range.

Table 3. Correlations Between Speech Recognition in Younger and Older Adults and Cognition and Learning.

	NFS 1		NFS 2 YA	SIN YA
	OA	YA		
Hearing	–0.50	–0.03	–0.04	0.1
TCS baseline	0.82	0.09	0.26	0.32
Vocabulary	0.47	0.30	0.40	0.35
Working memory	0.44	0.11	0.15	0.27
Block design	0.11	0.14	–0.09	0.27
Attention	–0.02	0.13	0.02	0.17
Slope	0.27	0.47	0.54	0.49

Note. Pearson correlations are shown. NFS 1 = natural-fast speech, Talker 1; NFS 2 = natural-fast speech, Talker 2; SIN = speech in noise; OA = older adults; YA = young adults; Hearing = average PTA; TCS baseline = average of first two sentences of time-compressed speech. Vocabulary, working memory, and block design—scaled scores from the corresponding tests; Attention = cost in the flanker task; Slope = rapid perceptual learning slope. Note that all correlations between rapid learning slopes and the cognitive variables were low ($r < .19$).

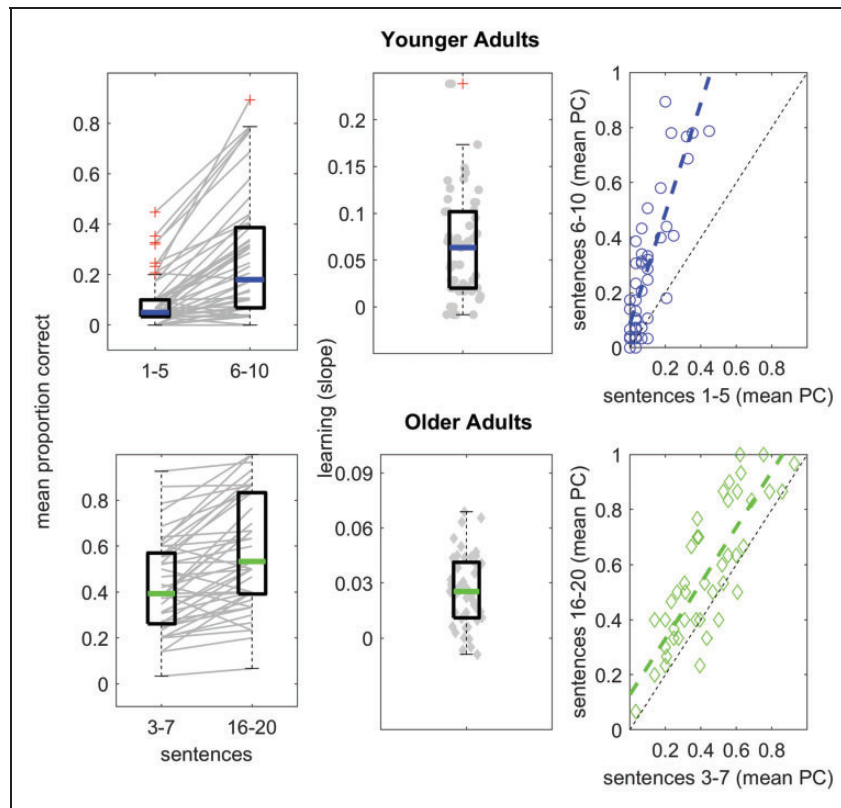


Figure 2. Rapid Learning of Time-Compressed Speech in Younger (Top Row) and Older (Bottom Row) Adults. The leftmost panel in each row shows performance (averaged over blocks of five sentences, not including baseline sentences; thin gray lines mark individual participants). Slopes of the learning curves (see text for details; individual data are shown with gray symbols) are shown in the middle panel of each row. Boxes mark the interquartile range (25th–75th percentile). The thick line within each box marks the median (blue for young adults and green for older adults). + signs mark outlying data points. The rightmost panel on each line shows performance on the final set of five sentences versus performance on the first set. Dashed diagonal lines mark $y = x$; thus, all symbols above this diagonal indicate learning. The dashed thicker lines (blue and green) show the linear fit between final and initial performance. Note that statistical analysis was based on the models described later, and fits are shown for demonstration only.

sentences in young adults and 20 sentences in older adults) in both older and younger adults. Speech recognition in older adults was evaluated with NFS of one talker and in young adults with NFS of two talkers and with SIN.

Modeling Speech Recognition as a Function of Hearing, Cognition, and Rapid Learning

To further account for the potential contribution of rapid learning to speech recognition, we modeled speech recognition as a function of rapid learning as well as other measures previously suggested in the literature to correlate with speech recognition (hearing thresholds, vocabulary, working memory, and attention). A series of generalized linear mixed models was ran using the lme4 package (Bates et al., 2015) in R (R Core Team, 2019). Random effects included intercepts for participant and sentence; predictors were scaled prior

to exporting the raw data to R. Following the recommendations of different sources about the analysis of proportions (e.g., Chen et al., 2017; Dunn & Smyth, 2018), we used binomial regressions with logistic link functions.

Three models were constructed for each index of speech recognition: A “null” model with hearing, vocabulary, working memory, and attention; a “baseline speech” model that included the same predictors of the “null” model as well as baseline recognition of TCS; and a “full” model that in addition to the predictors of the previous models also included the rapid learning slope. Because rapid learning in the current study was assessed using a speech recognition task, and different indices of speech recognition can be correlated (see Table 3), the “baseline speech” models were intended to account for the associations among different indices of speech recognition. The “null,” “baseline speech,” and “full” models were then compared to isolate the unique

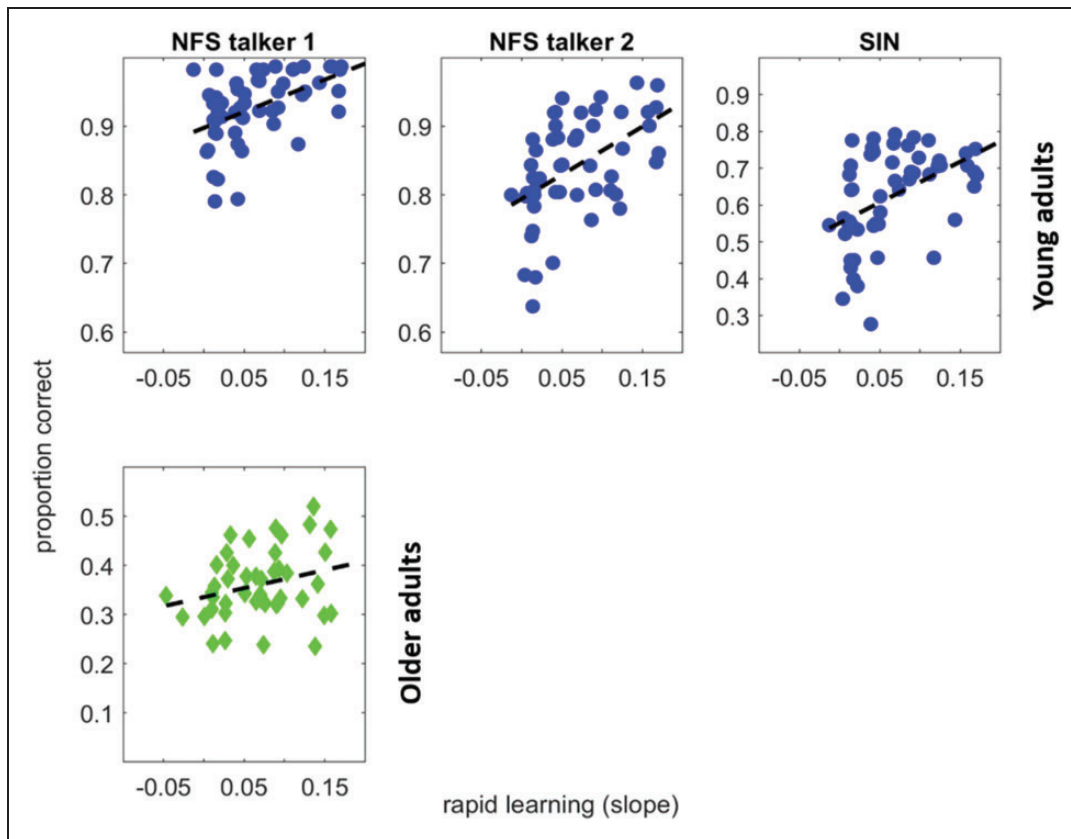


Figure 3. Speech Recognition Versus Rapid Learning. Proportions correct are plotted against the rapid learning slopes for younger (top row) and older (bottom row) adults. Left to right: natural-fast speech (NFS) produced by Talker 1, NFS produced by Talker 2, and speech-in-noise (SIN). Dashed lines show linear fits. Note that for the purpose of visual demonstration only, values of the learning slopes were adjusted to partial out the contribution of baseline recognition of time-compressed speech to the observed correlation (for details, see Manheim et al., 2018). Therefore, values on the x axis do not match the learning slopes shown in Figure 2.

Table 4. Estimates of Natural-Fast Speech Recognition (Talker 1) Based on the “Full” Model.

Predictor	Younger adults		Older adults	
	β (SE)	Z	β (SE)	Z
Hearing	0.0148 (0.0995)	0.15	-0.1568 (0.0520)	-3.01**
Vocabulary	0.1993 (0.0931)	2.14*	0.1910 (0.0585)	3.50**
Working memory	0.0386 (0.0937)	0.41	0.0725 (0.0497)	1.46
Attention	0.0584 (0.1076)	0.54	0.0747 (0.0469)	1.59
TCS baseline	-0.15 (0.1100)	-1.41	0.6498 (0.0578)	11.23***
Rapid learning	0.5170 (0.1183)	4.37***	0.2118 (0.0469)	4.52***

Note. TCS = time-compressed speech.

* $p < .05$. ** $p < .01$. *** $p < .001$.

contribution of learning. Given the differences in data collection and analysis described so far, we did not compare the models of older and younger adults directly.

Recognition of NFS. Table 4 shows the full model for NFS produced by Talker 1 for the two groups. In this model,

vocabulary and rapid learning were significant predictors of natural speech recognition in young adults. Rapid learning contributed about half a word for every 1 SD increase in learning. The “baseline speech” model in which hearing, cognition, and initial recognition of TCS were included was not significantly different

from the “null” model ($\chi^2 = 0.18$). On the other hand, the full model explained the data significantly better than the “baseline speech” model ($\chi^2 = 17.31$, $p < .001$), suggesting that rapid learning had a unique contribution to speech recognition, beyond that of the other variables.

In older adults, in the “full” model, hearing, vocabulary, initial TCS recognition, and rapid learning were all significant predictors of NFS recognition. In this group, model comparison suggests that the “baseline speech” model was significantly better than the null model ($\chi^2 = 52.89$, $p < .001$), but the full model was significantly better than the baseline speech model ($\chi^2 = 16.83$, $p < .001$). This suggests that while in older adults with hearing loss the strongest predictor (largest beta) of NFS recognition was recognition of a somewhat different type of distorted speech (TCS), learning was also a significant predictor, and the magnitude of its contribution was similar to those of hearing and vocabulary.

In young adults, modeling NFS of the second—slower but harder to recognize talker—resulted in similar outcomes, with vocabulary and rapid learning as significant predictors (see Table 5). Rapid learning improved recognition by about 6% for every 1 SD increase and vocabulary by about 4%. Model comparison suggested that the “baseline speech” model predicted the data better than the “null” model ($\chi^2 = 4.33$, $p < .05$); the “full” model that included rapid learning fitted the data better than the “baseline speech” model ($\chi^2 = 15.90$, $p < .001$).

Recognition of SIN in Young Adults. As shown in Table 5, rapid learning, vocabulary, and working memory were all significant predictors of SIN recognition in the full model. Rapid learning (despite being assessed with a different speech task) improved the recognition of SIN by approximately 0.4 words/1 SD increase in learning. Vocabulary and working memory improved recognition by more than 0.2 words/1 SD increase. The model that included baseline recognition of TCS fitted the data better than the model with only hearing and cognitive variables ($\chi^2 = 7.07$, $p < .01$), but the model that included learning

was a better fit than the “baseline speech” model ($\chi^2 = 14.86$, $p < .001$). Under the conditions of the current study, SIN was the hardest speech task for young adults. Therefore, this model demonstrates that dynamic factors such as working memory and learning may come to play a greater role in speech recognition when listeners find it particularly challenging.

Discussion

We investigated the contribution of rapid perceptual learning to speech recognition under adverse conditions. Even after accounting for hearing levels, baseline recognition of TCS, vocabulary, memory, and attention, rapid learning slopes remained significant predictors of speech recognition in both older and younger adults. Other factors also predicted the recognition of degraded speech: Vocabulary significantly predicted the recognition of NFS in the two groups. In older adults with age-related hearing loss, hearing and baseline recognition of TCS also related to the recognition of NFS. Working memory emerged as a significant predictor only for SIN (which was not tested in older adults). Although these findings are correlational and thus do not speak of causality, they at the very least fail to disprove the hypothesis that rapid perceptual learning is one of the factors that support speech recognition under adverse conditions. Two aspects of the findings are noteworthy: First, rapid perceptual learning of TCS was associated with the recognition of other types of speech (NFS and SIN), which in the case of young adults were all produced by different talkers. In both age groups, learning remained a significant predictor even after the correlations between recognition of NFS and TCS were accounted for. Therefore, while we cannot disentangle the causal direction from the current design, we can nevertheless conclude that prediction models that include learning better fit speech recognition data than models that do not include learning. Second, although we did not statistically compare prediction models between groups or speech conditions, the

Table 5. Estimates of Natural-Fast Speech (Talker 2) and Speech in Noise in Younger Adults Based on the “Full” Model.

Predictor	NFS (Talker 2)		SIN	
	β (SE)	Z	β (SE)	Z
Hearing	0.0159 (0.0778)	0.21	0.1475 (0.0921)	1.60
Vocabulary	0.2273 (0.0735)	3.09**	0.2589 (0.0877)	2.95**
Working memory	0.0667 (0.0735)	0.91	0.2388 (0.0880)	2.71**
Attention	-0.0242 (0.0805)	-0.30	0.1104 (0.0940)	1.21
TCS baseline	0.0534 (0.0858)	0.62	0.1248 (0.0982)	1.27
Rapid learning	0.3696 (0.0880)	4.20***	0.4080 (0.0984)	4.15***

Note. NFS = natural-fast speech; SIN = speech in noise; TCS = time-compressed speech.
** $p < 0.01$, *** $p < 0.001$.

structures of the best fitting models seem qualitatively different across groups and speech tasks (in young adults), consistent with other findings that different types of adverse conditions may result in the recruitment of different processes to support speech recognition (Bent et al., 2016; McLaughlin et al., 2018).

The major outcome of this study is that independent assessment of perceptual learning remains a significant predictor of perception in both younger and older adults on top of the variance explained by other cognitive processes and hearing thresholds. These findings replicate our previous observations that perceptual learning of TCS is associated with the recognition of NFS in younger and older adults with normal hearing as well as in older adults with age-related hearing loss (Manheim et al., 2018) and extend them to older adults with more severe age-related hearing loss. They also show that the contribution of learning is not entirely mediated by other processes that were already known to be predictive of speech recognition under adverse conditions (e.g., vocabulary).

In young adults, the current data further show that the association between rapid perceptual learning and the recognition of degraded speech might extend across different type of acoustic degradation. As discussed later, this association is consistent with the notion that rapid perceptual learning is an individual capacity that supports listening under challenging acoustic conditions. However, because we used the same talker to assess both rapid perceptual learning and SIN recognition, an alternative interpretation is that the association reflects talker familiarity. Previous studies indeed show that under challenging conditions, speech presented by familiar talkers is easier to recognize than speech presented by unfamiliar ones (e.g., Johnsrude et al., 2013; Nygaard & Pisoni, 1998). In the current study, listeners who were better able to learn the characteristics of the talker during the rapid learning phase were also better able to use those characteristics to support recognition of speech presented by the same talker in noise. Although this was not the case in a previous study in which the association between learning and the recognition of NFS did not depend on talker familiarity (Manheim et al., 2018), it could be that effects of talker familiarity require longer to emerge. Indeed, effects of talker familiarity were previously documented for highly familiar talkers such as spouses or college professors (Johnsrude et al., 2013; Newman & Evers, 2007) or after intensive training (Kreitewolf et al., 2017; Nygaard & Pisoni, 1998), whereas in the current study, previous experience with the familiar talker was limited to ten sentences.

That learning might be a capacity involved in speech perception under adverse conditions is consistent with the notion that dynamic processes (variously termed perceptual learning, adaptation, adjustment, and

recalibration) are recruited to support speech perception when the received signal is insufficient to automatically match existing lexical representations due to either listener or environmental challenges. This is by no means a new notion (Ahissar et al., 2009; Bidelman et al., 2019; Davis & Johnsrude, 2007; Guediche et al., 2014; Ronnberg et al., 2013), but our findings imply that this dynamic process could be an individual characteristic that is not specific to any single condition. Furthermore, given the across-listeners differences in learning, it seems that different listeners greatly differ in their ability to employ it in service of their speech recognition. Finally, findings such as ours may help bridge the gap between the hallmark specificity of perceptual learning (Eisner & McQueen, 2005; Green et al., 2018) and its potential role in speech perception under ecological conditions (Samuel & Kraljic, 2009). If learning is recruited online whenever adverse conditions are encountered, specificity should not matter for individuals with good learning, whereas individuals with poor learning will be at a constant disadvantage. Furthermore, if this is the case, this could be one of the reasons why speech recognition in some adverse conditions (such as the presence of competition from other talkers or babble noise) is so challenging, because an inspection of the learning data reported in previous studies suggest that perceptual learning in noise may be slower to emerge than the learning of highly TCS (Burk et al., 2006; Karawani et al., 2016; Schlueter et al., 2016).

Consistent with previous findings (Manheim et al., 2018; Schlueter et al., 2016), in the current study, learning was present but seemed reduced in individuals with age-related hearing loss, even though their initial performance was quite good. Specifically, we found that both learning slopes and effect sizes are numerically smaller in older adults despite them getting more favorable starting conditions and more trials to learn. Nevertheless, as in young adults, learning and vocabulary were still significant predictors of NFS recognition. Furthermore, despite having a restricted range of hearing deficits (see Figure 1), both hearing thresholds and baseline recognition of TCS were also significant predictors. The contribution of learning can thus be separated from the contribution of sensory and perceptual challenges to speech perception that are associated with age-related hearing loss (e.g., Benichov et al., 2012; Gordon-Salant & Fitzgibbons, 1997; Humes & Christopherson, 1991; Humes & Roberts, 1990; Pichora-Fuller et al., 1995), captured here by baseline recognition of TCS. Nevertheless, under environmental conditions, with no special allowances and no extra time, older adults with hearing loss are doubly challenged by both sensory and cognitive declines and declines in rapid learning. In the current data, the positive contributions of perceptual learning and vocabulary were similar in size to the negative contribution of hearing loss, perhaps providing

another reason for why individual older adults with similar audiograms are so different in their perceptual profiles. Listeners who maintain good learning into old age might be better able to compensate for adversity by employing rapid learning, whereas those with poor learning are less able to do so. Although tests in the current study were certainly audible, all listeners were tested in unaided conditions, in which sensory factors associated with the hearing loss could have been more dominant than if assessment had been conducted with hearing aids. Therefore, further studies are required to determine whether prediction models change when listeners are using their hearing aids and whether rapid learning is associated with adaptation to new hearing aids and/or cochlear implants. Further studies are also needed to determine why learning declines with age and whether the course of decline can be affected by various interventions.

The current outcomes about the associations between individual differences in speech recognition and cognitive and linguistic factors are generally in line with those of previous studies (Banks et al., 2015; Bent et al., 2016; McLaughlin et al., 2018). They show that different cognitive processes might have differential contributions to speech recognition under adverse conditions. Specifically, whereas young adults' vocabulary associated with the recognition of both NFS and speech-in-noise, working memory associated with the recognition of speech-in-noise only. Likewise, in older adults, vocabulary contributed to the account of individual differences in the recognition of NFS, but the contribution of the other cognitive variables was insignificant. Although the interpretation of insignificant outcomes is complicated, we are not the first to report lack of association between speech recognition on one hand and working memory or attention on the other. For example, in a study on individual differences in the recognition of different types of distorted speech, vocabulary consistently predicted performance across the different challenging conditions, whereas working memory predicted only some of the conditions (McLaughlin et al., 2018). Likewise, consistent with the current findings, flanker costs did not relate to the intelligibility of the distorted speech conditions used by Bent et al. (2016). The authors of these studies suggested that different abilities may help overcome particular types of deviation from canonical speech and that certain challenges may afford the recruitment of specific cognitive resources. Therefore, for the present findings, one possibility is that NFS was not sufficiently challenging to engage working memory in young adults (see Table 1), but this cannot explain why working memory and NFS recognition were not associated in older adults. Alternatively, the fast speech rates used in the current study could have resulted in a mismatch between the representation of the incoming stimuli and available

linguistic representations. According the Ease of Language Understanding Model, when such a mismatch occurs, the input fails to automatically activate existing long-term representations, and additional working memory resources are recruited in an attempt to resolve the mismatch (e.g., Ronnberg et al., 2019). However, rapid speech rates may have resulted in a mismatch that was too large to overcome, in which case recruiting working memory would have been unhelpful. In contrast, in noise the signal is masked, resulting in uncertainty about the correct match, in which case working memory could support speech recognition, leading to the correlation we observed.

Finally, this study had a number of limitations. First and foremost, our design makes it impossible to disentangle cause and effect, and it could be that poorer speech processing could interfere with perceptual learning. We nevertheless note that had this been the case, in the current study, larger learning slopes should have been observed in older and not in younger adults, because due to the levels of time compression, we selected young adults started out with substantially poorer recognition of TCS. Yet, despite poorer starting performance, young adults had larger learning slopes. Second, as explained earlier, there were a number of differences between the test conditions of older and younger adults which make a direct statistical comparison between the groups impossible. Therefore, additional studies in which a single prediction model can be attempted for all participants regardless of age and hearing status and with a larger range of challenging speech conditions are required. Third, sensory factors that could contribute to speech recognition (other than hearing thresholds) were not included in the current study. Instead, we relied on the baseline estimate of TCS recognition as a statistical control. As such factors could contribute to rapid auditory learning, they should be explored in future studies. Nevertheless, we believe these limitations are not sufficient to undermine the conclusion that rapid speech learning *could* be one of the factors contributing to individual variability in speech recognition under adverse conditions.

Acknowledgments

The authors thank Omri Gavish and Ayelet Nagar for their help.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this

article: The Israel Science Foundation (Grant 206/18) funded this work.

ORCID iD

Karen Banai  <https://orcid.org/0000-0002-2990-0470>

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. <https://doi.org/10.1121/1.3216914>
- Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, *25*(3), 736–740. <https://doi.org/10.1037/a0020054>
- 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. <https://doi.org/10.1098/rstb.2008.0253>
- Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2004). Aging: A switch from automatic to controlled processing of sounds? *Psychology and Aging*, *19*(1), 125–133. <https://doi.org/10.1037/0882-7974.19.1.125>
- Altmann, T. M., & Young, D. (1993, September). *Factors affecting adaptation to time-compressed speech* [Paper presentation]. Third European Conference on Speech Communication (EUROSPEECH '93) (pp. 333–336). Berlin, Germany.
- Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, *300*, 18–32. <https://doi.org/10.1016/j.heares.2013.03.006>
- 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–EL180. <https://doi.org/10.1121/1.4789864>
- Banai, K., & Lavner, L. (2020). Perceptual learning and speech perception: A new hypothesis. In A.A. Kressner, J. Regev, J. Christensen-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. (Eds.), *Proceedings of the International symposium on auditory and Audiological Research, Vol 7: Auditory learning in biological and artificial systems* (pp. 53–60). The Danavox Jubilee Foundation.
- 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. <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. *The Journal of the Acoustical Society of America*, *140*(3), 1686. <https://doi.org/10.1121/1.4962499>
- Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *Journal of the Acoustical Society of America*, *137*(4), 2015–2024. <https://doi.org/10.1121/1.4916265>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Benichov, J., Cox, L. C., Tun, P. A., & Wingfield, A. (2012). Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability, and cognitive function. *Ear and Hearing*, *33*(2), 250–256. <https://doi.org/10.1097/AUD.0b013e31822f680f>
- Bent, T., Baese-Berk, M., Borrie, S. A., & McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *Journal of the Acoustical Society of America*, *140*(5), 3775. <https://doi.org/10.1121/1.4966677>
- Bidelman, G. M., Price, C. N., Shen, D., Arnott, S. R., & Alain, C. (2019). Afferent-efferent connectivity between auditory brainstem and cortex accounts for poorer speech-in-noise comprehension in older adults. *Hearing Research*, *382*, 107795. <https://doi.org/10.1016/j.heares.2019.107795>
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. [https://doi.org/S0010-0277\(07\)00112-6](https://doi.org/S0010-0277(07)00112-6)
- Burk, M. H., Humes, L. E., Amos, N. E., & Strauser, L. E. (2006). Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear and Hearing*, *27*(3), 263–278. <https://doi.org/10.1097/01.aud.0000215980.21158.a2>
- Carbonell, K. M. (2017). Reliability of individual differences in degraded speech perception. *Journal of the Acoustical Society of America*, *142*(5), EL461. <https://doi.org/10.1121/1.5010148>
- Chen, K., Cheng, Y., Berkout, O., & Lindhiem, O. (2017). Analyzing proportion scores as outcomes for prevention trials: A statistical primer. *Prevention Science*, *18*, 312–321. <https://doi.org/10.1007/s11121-016-0643-6>
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.
- Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, *47*(Pt 3), 207–239.
- Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *Journal of Speech, Language, and Hearing Research*, *61*(8), 1855–1874. https://doi.org/10.1044/2018_JSLHR-S-17-0392
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147. <https://doi.org/10.1016/j.heares.2007.01.014>
- 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. <https://doi.org/2005-04168-006>
- DeCaro, R., Peelle, J. E., Grossman, M., & Wingfield, A. (2016). The two sides of sensory-cognitive interactions: Effects of age, hearing acuity, and working memory span

- on sentence comprehension. *Frontiers in Psychology*, 7, 236. <https://doi.org/10.3389/fpsyg.2016.00236>
- Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America*, 76(1), 87–96.
- Dunn, P. K., & Smyth G.K. (2018). *Generalized linear models with examples in R, chapter 9*. Springer. <https://doi.org/10.1007/978-1-4419-0118-7>
- 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.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, 67(2), 224–238. <https://doi.org/10.3758/Bf03206487>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, 16(1), 143–149.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *Journal of the Acoustical Society of America*, 121(3), 1701–1708. <https://doi.org/10.1121/1.2436635>
- Gordon-Salant, S., & Fitzgibbons, P. J. (1995). Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index. *Journal of Speech and Hearing Research*, 38(3), 706–713. <https://doi.org/10.1044/jshr.3803.706>
- Gordon-Salant, S., & Fitzgibbons, P. J. (1997). Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language, and Hearing Research*, 40(2), 423–431. <https://doi.org/10.1044/jslhr.4002.423>
- Gordon-Salant, S., & Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research*, 44(4), 709–719. <https://doi.org/10.1044/1092-43882001/056>
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Schurman, J. (2010). Short-term adaptation to accented English by younger and older adults. *Journal of the Acoustical Society of America*, 128(4), EL200–EL204. <https://doi.org/10.1121/1.3486199>
- Green, C. S., Banai, K., Lu, Z. L., & Bavelier, D. (2018). Perceptual learning. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience (Vol. 2, pp. 1–47)*. John Wiley & Sons, Inc.
- 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.
- Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, 7, 126. <https://doi.org/10.3389/fnsys.2013.00126>
- Humes, L. E., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. *Journal of Speech and Hearing Research*, 34(3), 686–693.
- Humes, L. E., & Dubno, J. R. (2010). Factors affecting speech understanding in older adults. In S. Gordon-Salant, R. D. Frisina, R. R. Fay, & A. Popper (Eds.), *The aging auditory system* (pp. 211–257). Springer.
- Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., Gates, G. A., Wilson, R. H., & Wingfield, A. (2012). Central presbycusis: A review and evaluation of the evidence. *Journal of the American Academy of Audiology*, 23(8), 635–666. <https://doi.org/10.3766/jaaa.23.8.5>
- Humes, L. E., & Roberts, L. (1990). Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research*, 33(4), 726–735.
- Janse, E. (2009). Processing of fast speech by elderly listeners. *Journal of the Acoustical Society of America*, 125(4), 2361–2373. <https://doi.org/10.1121/1.3082117>
- Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, 24(10), 1995–2004.
- Karawani, H., Bitan, T., Attias, J., & Banai, K. (2016). Auditory perceptual learning in adults with and without age-related hearing loss. *Frontiers in Psychology*, 6, 2066. <https://doi.org/10.3389/fpsyg.2015.02066>
- Karawani, H., Lavie, L., & Banai, K. (2017). Short-term auditory learning in older and younger adults. In S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjaerg, T. Andersen, & T. Poulsen. (Eds.), *Proceedings of the International Symposium on Auditory and Audiological Research, Vol 6. Adaptive Processes in Hearing* (pp. 1–8). The Danavox Jubilee Foundation.
- Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit talker training improves comprehension of auditory speech in noise. *Frontiers in Psychology*, 8, 1584.
- Letowski, T., & Poch, N. (1996). Comprehension of time-compressed speech: Effects of age and speech complexity. *Journal of the American Academy of Audiology*, 7(6), 447–457.
- Manheim, M., Lavie, L., & Banai, K. (2018). Age, hearing, and the perceptual learning of rapid speech. *Trends in Hearing*, 22, 2331216518778651. <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. <https://doi.org/10.1016/j.cogpsych.2009.04.001>
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech.

- Attention, Perception, & Psychophysics*, 80(6), 1559–1570. <https://doi.org/10.3758/s13414-018-1537-4>
- Nagaraj, N. K. (2017). Working memory and speech comprehension in older adults with hearing impairment. *Journal of Speech Language and Hearing Research*, 60(10), 2949–2964. https://doi.org/10.1044/2017_JSLHR-H-17-0022
- 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. <https://doi.org/10.1371/journal.pbio.0060126>
- Neger, T. M., Rietveld, T., & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in Human Neuroscience*, 8, 628. <https://doi.org/10.3389/fnhum.2014.00628>
- Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, 35(1), 85–103.
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376.
- Oberfeld, D., & Klockner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife*, 5, e16747. <https://doi.org/10.7554/eLife.16747>
- 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. <https://doi.org/10.1037/0096-1523.31.6.1315>
- Peelle, J. E., & Wingfield, A. (2016). The neural consequences of age-related hearing loss. *Trends in Neurosciences*, 39(7), 486–497. <https://doi.org/10.1016/j.tins.2016.05.001>
- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, 97(1), 593–608. <https://doi.org/10.1121/1.412282>
- Prior, A., & Bentin, S. (2006). Differential integration efforts of mandatory and optional sentence constituents. *Psychophysiology*, 43(5), 440–449. <https://doi.org/10.1111/j.1469-8986.2006.00426.x>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rönnerberg, J., Holmer, E., & Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *International Journal of Audiology*, 58(5), 247–261. <https://doi.org/10.1080/14992027.2018.1551631>
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, 31. <https://doi.org/10.3389/fnsys.2013.00031>
- Rotman, T., Lavie, L., & Banai, K. (2020). Rapid perceptual learning of time-compressed speech and the perception of natural fast speech in older adults with presbycusis. In A.A. Kressner, J. Regev, J. Christensen-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. (Eds.), *Proceedings of the International Symposium on Auditory and Audiological Research, Vol 7: Auditory learning in biological and artificial systems* (pp. 93–100). The Danavox Jubilee Foundation.
- Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, 62, 49–72. <https://doi.org/10.1146/annurev.psych.121208.131643>
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207–1218. <https://doi.org/10.3758/APP.71.6.1207>
- Scharenborg, O., & Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, & Psychophysics*, 75(3), 525–536. <https://doi.org/10.3758/s13414-013-0422-4>
- Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, perception & psychophysics*, 77(2), 493–507. <https://doi.org/10.3758/s13414-014-0792-2>
- Schlueter, A., Lemke, U., Kollmeier, B., & Holube, I. (2016). Normal and time-compressed speech: How does learning affect speech recognition thresholds in noise? *Trends in Hearing*, 20, 2331216516669889. <https://doi.org/10.1177/2331216516669889>
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *Journal of the Acoustical Society of America*, 123(1), 489–499. <https://doi.org/10.1121/1.2783762>
- Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences of the United States of America*, 113(12), E1747–E1756. <https://doi.org/10.1073/pnas.1523266113>
- Sommers, M. S., Spehar, B., Tye-Murray, N., Myerson, J., & Hale, S. (2020). Age differences in the effects of speaking rate on auditory, visual, and auditory-visual speech perception. *Ear and Hearing*, 41, 549–560. <https://doi.org/10.1097/AUD.0000000000000776>
- Souza, P., & Arehart, K. (2015). Robust relationship between reading span and speech recognition in noise. *International Journal of Audiology*, 54(10), 705–713. <https://doi.org/10.3109/14992027.2015.1043062>
- 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. <https://doi.org/10.1044/1092-4388.2008.038>
- Theodore, R. M., Monto, N. R., & Graham, S. (2019). Individual differences in distributional learning for speech: What's ideal for ideal observers? *Journal of Speech, Language, and Hearing Research*, 63(1), 1–13. https://doi.org/10.1044/2019_JSLHR-S-19-0152
- Tierney, A., Rosen, S., & Dick, F. (2019). Speech-in-speech perception, nonverbal selective attention, and musical

- training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(5), 968–979. <https://doi.org/10.1037/xlm0000767>
- Verhelst, W., & Roelands, M. (1993). *An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech* [Paper presentation]. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Minneapolis, MN, United States.
- Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology*, 40(5), 579–585.
- Wingfield, A., & Tun, P. A. (2001). Spoken language comprehension in older adults: Interactions between sensory and cognitive change in normal aging. *Seminars in Hearing*, 22(3), 287–302.
- Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*, 14(3), 144–148.