

# Working memory training to improve speech perception in noise across languages

Erin M. Ingvalson<sup>a)</sup>

*School of Communication Science and Disorders, Florida State University, 201 West Bloxham Street, Tallahassee, Florida 32306, USA*

Sumitrajit Dhar

*Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA*

Patrick C. M. Wong

*Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Room G03, Leung Kau Kui Building, Shatin, N. T., Hong Kong Special Administrative Region*

Hanjun Liu

*Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, China*

(Received 20 October 2014; revised 4 May 2015; accepted 11 May 2015)

Working memory capacity has been linked to performance on many higher cognitive tasks, including the ability to perceive speech in noise. Current efforts to train working memory have demonstrated that working memory performance can be improved, suggesting that working memory training may lead to improved speech perception in noise. A further advantage of working memory training to improve speech perception in noise is that working memory training materials are often simple, such as letters or digits, making them easily translatable across languages. The current effort tested the hypothesis that working memory training would be associated with improved speech perception in noise and that materials would easily translate across languages. Native Mandarin Chinese and native English speakers completed ten days of reversed digit span training. Reading span and speech perception in noise both significantly improved following training, whereas untrained controls showed no gains. These data suggest that working memory training may be used to improve listeners' speech perception in noise and that the materials may be quickly adapted to a wide variety of listeners. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4921601>]

[ICB]

Pages: 3477–3486

## I. INTRODUCTION

Recent years have seen a surge of interest in cognitive training. Among other claims, improving one's cognitive performance has been suggested to lead to improved academic performance (Klingberg, 2010), improved language learning (Ingvalson and Wong, 2013), and reduced incidence of dementia (Willis *et al.*, 2006).

Much of the interest in cognitive training has focused on improving working memory. Training working memory has received particular attention because larger working memory capacities have been linked to better academic performance, better language learning, and reduced incidence of pathological aging (Klingberg, 2010; Morrison and Chein, 2010). Additionally, it has become apparent that working memory plays a role in listeners' ability to perceive speech in noisy situations. For both individuals with normal hearing and individuals who use hearing aids, listeners with larger working memory capacities appear to have more success perceiving speech in noise. Working memory capacity successfully predicts speech recognition in noise by normal hearing older

adults (Gordon-Salant and Fitzgibbons, 1997; Parbery-Clark *et al.*, 2009). In older adults who use hearing aids, greater working memory capacity also predicts more success recognizing speech in noise (Foo *et al.*, 2007; Lunner and Sundewall-Thorén, 2007). Within an individual listener, working memory capacity is correlated with degree of listening effort and success perceiving speech in noise (Koelewijn *et al.*, 2014; Zekveld *et al.*, 2011).

Increasingly frequent complaints of speech perception difficulties in background noise have drawn attention to the relationship between working memory and speech perception in noise. Listeners with hearing aids, users of cochlear implants, as well as older adults with normal hearing, all complain of difficulty perceiving speech in noisy situations. As the population ages (Kinsella and He, 2008), more individuals are likely to suffer from difficulties in speech perception in noise, making effective interventions increasingly necessary. Knowing that working memory capacity predicts speech perception in noise success suggests that one avenue for intervention may be cognitive training: increasing listeners' working memory capacity may lead to improved speech perception in noise performance.

Current knowledge supports the hypothesis that working memory capacity can be improved. Older adults with no

<sup>a)</sup>Electronic mail: erin.ingvalson@cci.fsu.edu

known cognitive impairment show increases in working memory capacity following working memory training (Bherer *et al.*, 2005; Li *et al.*, 2008). More generalized cognitive training, which emphasizes working memory capacity, speed of processing, inhibitory control, and semantic memory has also been shown to improve working memory capacity in healthy older adults (Ball *et al.*, 2002). Older adults with a diagnosis of mild cognitive impairment, a precursor to dementia, also show a benefit of cognitive training. Following training, older adults with mild cognitive impairment showed gains not only on measures of working memory but also on measures of cognitive health such as the Mini Mental State Exam (Folstein *et al.*, 1975) and the Dementia Rating Scale (O'Bryant *et al.*, 2008).

However, though the available data support the claim that working memory capacity can be increased in older adults, it is less clear whether this increase would lead to improvements in speech perception in noise. The extent to which cognitive training gains transfer to tasks that are only distally related to the training, called far transfer, remains an open question. Some studies have found success with far transfer to skills such as inhibitory control (Persson and Reuter-Lorenz, 2008) or reading comprehension (Chein and Morrison, 2010). On the other hand, other studies have found little evidence of far transfer following working memory training (Dahlin *et al.*, 2008; Schmiedek *et al.*, 2010). A direct comparison of far transfer in younger and older adults suggested that only younger adults may benefit from far transfer (Schmiedek *et al.*, 2010). Supporting the potential weakness of far transfer in older adulthood, both a review and a meta-analysis of cognitive training in older adulthood found little evidence of benefit in distally related tasks post-training (Morrison and Chein, 2010). A recent meta-analysis of working memory training, although not specific to older adults, also found little evidence of far transfer across the trained populations (Melby-Lervåg and Hulme, 2013). Examining the lists of tasks included, however, reveals that assessments of far transfer in older adults often include tasks that, like working memory, decline with age [e.g., fluency tasks (Dahlin *et al.*, 2008)] or that have been shown to share neural substrates with working memory [e.g., episodic memory tasks (Cabeza *et al.*, 2002)] but not where any behavioral connection has been made between working memory performance and performance on the tested task. There is therefore little theoretical reason to expect increases in working memory to lead to increases in these tasks, other than an expectation that training working memory should lead to across-the-board increases in cognitive ability (Shipstead *et al.*, 2012). Evidence of far transfer is apparent where links have been made between greater working memory capacity and greater performance [e.g., reading comprehension; (Morrison and Chein, 2010)]. The mismatch between what can theoretically be expected to improve following training and the distal outcome measures was also noted in a recent review of auditory training (Henshaw and Ferguson, 2013). Henshaw and Ferguson concluded that more research into auditory training is needed and that assessments of auditory training should include measures that will capture a clinically significant improvement in listening ability. Looking

over the data in these reviews, it is apparent that both working memory and auditory training studies would benefit from more theoretically driven outcome measures, which will provide a reliable indicator of whether far transfer is possible. In the present study, we note that behavioral connections have been made between working memory capacity and speech perception and noise (Foo *et al.*, 2007; Lunner, 2003; Lunner and Sundewall-Thorén, 2007) leading to a theoretical connection between the two (Rönnberg *et al.*, 2008). We therefore hypothesize that improvements in working memory will produce far transfer and lead to improvements in speech perception in noisy situations.

A secondary goal of working memory training to improve speech perception in noise is its potential transferability to additional language environments. Commercially available speech perception in noise training programs such as LACE or SPATS (Miller *et al.*, 2008; Sweetow and Sabes, 2006) are successful in improving speech perception in noise, but they rely extensively on sentence-level speech items. Currently these materials are only available in American English, limiting their utility to listeners who are not proficient in English or who do not speak American English as their native dialect (Brouwer *et al.*, 2012; Van Engen and Bradlow, 2007). Conversely many working memory training programs rely on very simple items, such as digits, letters, or single-syllable words (Klingberg, 2010; Morrison and Chein, 2010). Although sentence-level materials would take a large amount of effort to translate across languages, necessitating checks to ensure items are grammatically correct, culturally relevant, and produced by native speakers (Brouwer *et al.*, 2012; Wong *et al.*, 2008), translating typical working memory items across languages requires only production by native speakers. Furthermore, digits, letters, and single words are very ecologically valid in that they are readily familiar to listeners with hearing impairments and do not require a high semantic load and thus are relatively robust against hearing loss (Smits *et al.*, 2013; Wilson *et al.*, 2010), ensuring the listener is engaging in a working memory task and not a speech perception task. We therefore hypothesize that working memory training will demonstrate speech-in-noise benefits in multiple languages.

In addition to the ecological validity of digits, letters, and single words for speech in noise testing and training (Smits *et al.*, 2013; Wilson *et al.*, 2010), there is the concern that sentence-level auditory materials may not be optimal for improving working memory. Traditionally, training working memory has been thought of as attempting to increase the capacity of the central executive (Baddeley, 2003; Klingberg, 2010). In commercially available products, listeners are asked to guess what the missing word in a sentence might be or to indicate which word appeared before a target word in a sentence. These tasks rely heavily on the ability to extract meaning from the sentence. Though extracting meaning is important for understanding speech in noise, it is not a working memory task in itself. Instead the task of extracting meaning—and thereby filling in missing words or finding target words—requires interactions of working memory, lexical access, and access of long-term memory (Hickok and Poeppel, 2007). Improvement may

therefore be a result of listeners' increased practice filling in information missed in noise than true working memory gains. Here we aim to improve listeners' working memory performance using the simple stimuli customary among working-memory training paradigms. Simultaneously, our training paradigm provides practice listening to speech in a variety of noise types, including steady-state and environmental noise at two signal-to-noise ratios (SNRs). The speech-perception-in-noise practice was introduced to avoid asymptotic performance on the working memory task, but it provides the additional benefit of allowing listeners to practice perceiving speech in noise while increasing auditory working memory capacity. We differentiate the practice perceiving speech in noise in earlier studies from the present study by noting that those studies used whole sentences, whereas we use only digits. Although the recognition of whole sentences in noise may utilize multiple memory systems, we suggest that rehearsing and storing a string of digits will utilize only working memory (specifically, the phonological loop; [Baddeley, 1986](#)). Consequently, we hypothesize that the training paradigm will be associated with gains on both working memory measures and measures of speech perception in noise.

When developing an intervention for speech perception in noise, one consideration is the outcome measure. This consideration is compounded when the intervention is intended to be translated across multiple languages, as is the case here. In the present study, we opted to assess speech-perception-in-noise gains in Mandarin via the M-HINT and in English via the QuickSIN. The QuickSIN has repeatedly been shown to be more sensitive than the HINT as a measure of speech perception in noise ([Duncan and Aarts, 2006](#); [Wilson et al., 2007](#)) particularly for listeners with normal hearing ([Parbery-Clark et al., 2009](#)). Ideally, therefore, we would use the QuickSIN for assessment in both language environments, but the M-HINT is the only speech-perception-in-noise assessment currently available in Mandarin Chinese ([Wong et al., 2008](#)). We therefore opted to use the more sensitive test when available to better capture how working memory training might lead to speech perception in noise gains. We note that performance on the HINT and performance on the QuickSIN are closely correlated ([Duncan and Aarts, 2006](#); [Weber et al., 2010](#)), suggesting that gains seen on one test are likely to also be seen on the other and to be of similar magnitude ([Weber et al., 2010](#)).

The present experiment trained working memory using digits. Based on the demonstrated relationship between increased working memory capacity and speech perception in noise, we hypothesize that gains in training will correspond to working memory and speech-perception-in-noise gains. We further hypothesize that digits will be easily translated across languages and that we will see gains in more than one linguistic context. We completed two experiments, one in Mandarin Chinese and one in English. We conducted two experiments, in two language environments, to test our hypothesis that digits would readily transfer to multiple languages and correspond to speech-in-noise gains across languages. In both experiments, we found that backward

digit-span training was associated with speech-perception-in-noise gains.

## II. EXPERIMENT 1

### A. Methods

#### 1. Participants

Twenty-five native Mandarin Chinese speakers (15 female,  $22.08 \pm 1.93$  yr, mean  $\pm$  SD) were recruited and run at Sun-Yat Sen University of China. Participants self-reported normal hearing and had no known cognitive deficits. Fifteen participants (eight female) were assigned to the training group; ten (seven female) were assigned to the control group. Group assignments were random; trained listeners were simultaneously enrolled in an additional study of speech perception in noise that used identical protocols but that required a larger number of subjects, resulting in the disparity in group size. There were no differences between the trained and control participants in gender, level of education, or age, all  $p > 0.05$ .

#### 2. Materials

A native male Mandarin speaker recorded the digits one through nine. Recordings were sampled at 44 100 Hz and 16 bit accuracy then RMS matched to a 70 dB sound pressure level (SPL) pure tone at 1 kHz. Two types of noise were added to the recordings to create two distinct speech-in-noise environments. We created steady-state noise shaped to match the long-term average spectrum of the recordings using MATLAB. The steady-state noise was mixed with the target recordings in ADOBE AUDITION to create stimuli with SNR of  $-5$  and  $-10$  dB. Pilot testing indicated these SNR levels were challenging but sufficiently perceptible to not interfere with digit rehearsal for the working memory task.

We obtained a set of 27 non-speech sounds (e.g., mechanical, human non-speech, animal, and musical) from online databases. Non-speech distractors were chosen because they serve as a fluctuating masker and have cross-linguistic informational masking value ([Garcia Lecumberri and Cooke, 2006](#)). Although the informational masking of non-speech sounds is not likely to be as great as with speech, using speech as a masker would require the development of new maskers for each language environment; non-speech maskers thereby allow for more rapid transfer across languages. Twelve sounds were animal sounds (e.g., bird song, dog bark), two sounds were mechanical (e.g., clock ticking, gun shot), five sounds were non-speech human (e.g., cough, laugh), and eight sounds were musical (e.g., violin, flute). These sounds were cropped to 1 s, RMS matched in amplitude to the calibration tone, then mixed with the target recordings at  $-5$  and  $-10$  dB SNR. All non-speech distractors were mixed with all targets at all SNR levels, for a total of 468 stimulus combinations.

In both the steady-state and non-speech noise conditions, total stimulus duration was 1 s. Targets were placed 15 ms after the onset of the distractor; targets offset 15 ms prior to stimulus offset. Presentation level for the final stimulus was set to 70 dB SPL.

Speech perception in noise performance was assessed using the Mandarin Chinese HINT (Wong *et al.*, 2008). Performance on the HINT is measured in reception threshold for speech (RTS). Working memory performance was assessed using a Chinese version of the reading span test (Daneman and Carpenter, 1980). Both the reading span task used for assessment and the backward digit span task used for training are considered complex working memory tasks (Daneman and Merikle, 1996), suggesting they tap the same working memory mechanisms, and training of one task is expected to thereby transfer to the other (Baddeley, 2003; Chein and Morrison, 2010).

### 3. Procedure

Total experimental duration was 12 days. Participants assigned to the control group completed the pre- and post-tests on the 1st and 12th day, and made no contact with the experimenters on the intervening days. Participants assigned to the training group trained on the 2nd through 11th days.

*a. Testing.* Before testing, audibility of the speech in noise was verified for all listeners. List 1 of the M-HINT was presented in quiet. Following correct identification of the list, participants indicated their ability to perceive the M-HINT noise in isolation. Loudness levels for both targets and speech were adjusted until listeners correctly identified all target words in quiet and indicated perception of the noise. All participants were able to perceive both the target and the noise at the recommended presentation level of 65 dB A.

At the pre- and post-test participants completed the M-HINT and the reading span test. Three M-HINT lists were chosen at random for each test presentation with the constraint that different lists be used at pre- and post-test within subject. All lists were presented via Etymotic ER1 insert earphones in a sound-attenuated booth. An experimenter outside the booth scored keywords repeated correctly according to standard procedure (Wong *et al.*, 2008). Prior to completing the three test lists, participants heard and repeated one practice list to ensure audibility and familiarize participants with the task.

The reading span test was given and scored according to standard procedures (Daneman and Carpenter, 1980). Participants read a sentence, indicated whether the sentence made semantic sense or not, then remembered the final word in the sentence. After reading all the sentences in a list, participants recalled all the final words from that list. There were three lists of the same length per block. Reading span was defined as the last block on which the participant correctly recalled all the items on two of three lists.

*b. Training.* Training was implemented in E-Prime (Psychological Testing Services, Pittsburgh, PA) using a backward digit span paradigm (i.e., if the participant heard, “1-2-3-4,” the correct response was, “4-3-2-1”). On each trial, participants heard a list of digits presented over Sennheiser HD-250 headphones in a sound-attenuated booth. At the end of the list, the participant was visually cued by

the computer to respond. No feedback was given on individual trials.

We developed an adaptive training metric that adjusted the length of digit span according to listener ability. There were seven digit lists per block. If block accuracy was greater than 70%, span length on the subsequent block was increased by one digit. Conversely, if block accuracy was less than 50%, span length on the subsequent block was decreased by one digit. Ten blocks were completed each day. Each day’s first span length was the same as the last span length of the previous training day; span length on the first block of the first day was set to 2.

Pilot testing demonstrated that participants quickly reached asymptotic performance when performing the digit span task in quiet. We therefore implemented training in noise to maintain task engagement. On training days one and two, participants trained in quiet. Training days three and four were done in steady-state noise at  $-5$  dB SNR. Training days five and six were done in non-speech distractors at  $-5$  dB SNR; distractors were chosen at random for each digit presentation by the training program. Days seven through ten replicated training days three through six, but at  $-10$  dB SNR.

### B. Results

Figure 1 demonstrates the training gains. Participants made consistent increases in their digit spans throughout training. An analysis of variance (ANOVA) for linear trends was significant,  $F(1,8) = 379.15$ ,  $p < 0.001$ , confirming the upward trajectory.

We entered the data from the M-HINT into a 2 (group: trained vs control)  $\times$  2 (session: pre- vs post-test) mixed ANOVA where session was the within-subjects factor. We found the expected interaction of group and session,  $F(1,23) = 11.24$ ,  $p = 0.003$ , shown in Fig. 2. Planned  $t$ -tests showed a significant improvement by the trained subjects,  $t(14) = 5.87$ ,  $p < 0.001$  (pre-test RTS  $M = -3.01$  dB SNR, post-test RTS  $M = -3.85$  dB SNR), but not the control subjects,  $t(9) = 0.34$ ,  $p = 0.74$  (pre-test RTS  $M = -2.81$  dB

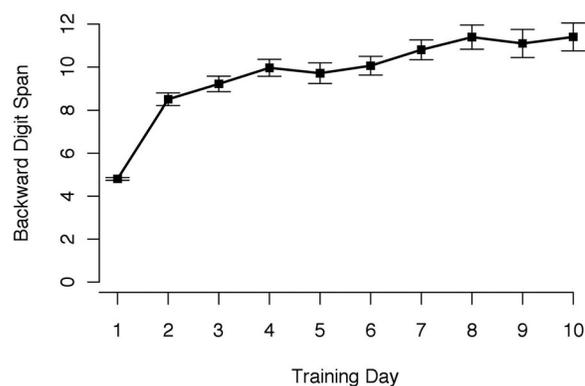


FIG. 1. Mean backward digit span performance for the native Mandarin speakers as a function of training day. Training days 3, 4, 7, and 8 were presented in speech-shaped noise, and days 5, 6, 9, and 10 were presented in multiple distractors (e.g., nature sounds, music). Training days 3–6 were presented at  $-5$  dB SNR and days 7–10 were at  $-10$  dB SNR. Error bars are standard error of the mean.

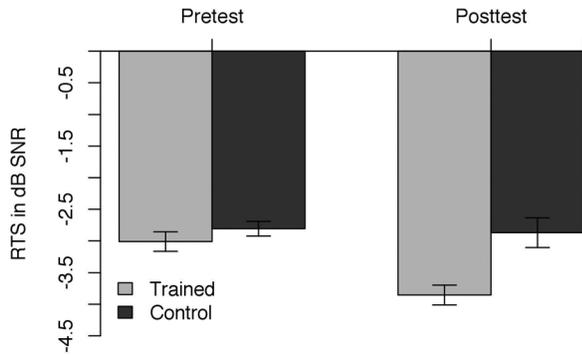


FIG. 2. Mean dB RTS performance on the HINT at pre- and post-test by the trained and control native Mandarin speaking subjects. Trained subjects showed a significant improvement on their ability to perceive speech in noise whereas control subjects did not. Error bars are standard error of the mean.

SNR, post-test RTS  $M = -2.87$  dB SNR). There was also a main effect of group,  $F(1,23) = 7.65$ ,  $p = 0.01$  and one of session,  $F(1,23) = 21.69$ ,  $p < 0.001$ . Overall the trained group performed better than the control group (trained RTS  $M = -3.43$  dB SNR, control RTS  $M = -2.84$  dB SNR) and overall performance was better at the post-test than at the pre-test (pre-test RTS  $M = -2.93$  dB SNR, post-test RTS  $M = -3.46$  dB SNR). However, these main effects should be interpreted with an eye to the significant interaction.

The reading span data were entered into a separate 2 (group: trained vs control)  $\times$  2 (session: pre- vs post-test) mixed ANOVA. We again found the anticipated interaction of group and session,  $F(1,23) = 7.87$ ,  $p = 0.01$ , shown in Fig. 3. Planned  $t$ -tests again revealed a significant improvement by the trained group,  $t(14) = 3.56$ ,  $p = 0.003$  (pre-test  $M = 4.87$  words, post-test  $M = 6.33$  words), but not the control group,  $t(9) = 0$ ,  $p = 1$  (pre-test  $M = 4.80$  words, post-test  $M = 4.80$  words). The main effect of session was also significant,  $F(1,23) = 11.80$ ,  $p = 0.002$ . Post-test performance was better than pretest performance overall (pre-test  $M = 4.84$  words, post-test  $M = 5.72$  words), but, again, this should be interpreted with an eye to the significant interaction.

### C. Discussion

The results from the first experiment indicate a benefit of working memory training for improving speech

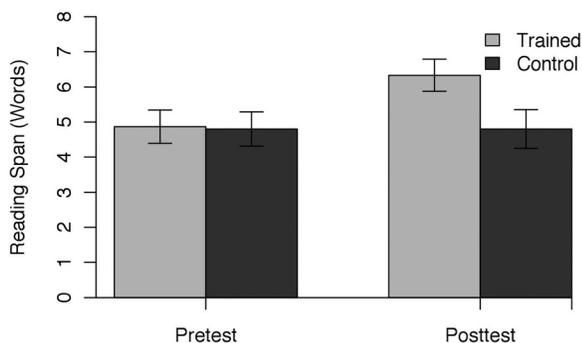


FIG. 3. Mean reading span performance by the native Mandarin speaking subjects. Trained subjects significantly increased their reading spans following training, whereas there was no change in control spans at posttest. Error bars are standard error of the mean.

perception in noise. Trained listeners showed gains in span length throughout training, gains of approximately 1.5 words on the reading span test and gains of 0.85 dB in their reception thresholds for sentences (RTS). The fact that these gains were seen even in young, cognitively healthy, normal-hearing listeners is especially impressive as this population typically does not show difficulties perceiving speech in noisy situations nor deficits in their working memory performance.

Despite the significant gains seen on all components, this experiment does not address the fact that currently available commercial speech perception in noise training paradigms are difficult to translate across languages. One of the aims in using digits as training stimuli is the hypothesis that digits are easily translatable across languages, making the training paradigm easily adaptable to improve both working memory and speech perception in noise for listeners in a variety of languages. Experiment 2 was designed to test this hypothesis.

## III. EXPERIMENT 2

### A. Methods

#### 1. Participants

Twenty-three native English speakers (15 female,  $19.91 \pm 1.34$  yr) were recruited and run at Northwestern University. All participants had hearing thresholds of 25 dB hearing level (HL) or better, obtained in a sound-attenuated booth and reported no known cognitive deficits. Twelve participants (7 female) were assigned to the trained group; 11 (8 female) were assigned to the control group. Group assignment was random. There was no significant difference between the trained and control groups on gender, age, or educational status, all  $p > 0.05$ .

#### 2. Materials

A native English speaker recorded the digits 1–9. Recording conditions, stimulus processing, and distractors were identical to experiment 1. Speech perception in noise testing was administered using the Quick Speech in Noise Test (Killion *et al.*, 2004). Performance on the QuickSIN is measured in SNR loss or the SNR required to perceive speech in noise compared to normal performance (Killion *et al.*, 2004). Working memory performance was assessed using the reading span test (Daneman and Carpenter, 1980).

#### 3. Procedure

When administering the QuickSIN, as with the M-HINT, listeners were given three randomly selected lists at each test session with the caveat that lists could not be repeated across sessions within listener. Lists were administered according to standard procedures. Audibility of targets and noise were verified using two practice lists according to standard procedures (Killion *et al.*, 2004); no listeners required a change from standard loudness levels (70 dB HL).

Lists were selected from the subset of lists known to be equivalent (McArdle and Wilson, 2006).

All other testing and training procedures were identical to experiment 1.

## B. Results

Figure 4 demonstrates training gains. Visually comparing training gains by the Mandarin listeners (Fig. 1) to the training gains by the English listeners (Fig. 4) suggests that the English listeners struggled more with speech-shaped noise at  $-10$  dB SNR than did the Mandarin listeners. Nonetheless an ANOVA for linear trends indicates that gains were significant over the entire training period,  $F(1,8) = 81.73, p < 0.001$ .

We entered the data obtained from the QuickSIN into a 2 (group: trained vs control)  $\times$  2 (session: pre- vs post-test) mixed ANOVA where session was the within-subjects factor. We found the expected interaction of group and session,  $F(1,21) = 7.76, p = 0.01$ , shown in Fig. 5. Planned  $t$ -tests showed a significant improvement by the trained subjects,  $t(11) = 2.82, p = 0.02$  (pre-test  $M = -0.33$  dB SNR loss, post-test  $M = -1.28$  dB SNR loss) but not the control subjects,  $t(10) = 1.18, p = 0.26$  (pre-test  $M = -0.5$  dB SNR loss, post-test  $M = -0.08$  dB SNR loss). None of the main effects were significant, group:  $F(1,21) = 1.21, p = 0.28$ ; session:  $F(1,21) = 1.40, p = 0.25$ .

The reading span data were entered into a separate 2 (group: trained vs control)  $\times$  2 (session: pre- vs post-test) mixed ANOVA. We again found the expected interaction of group and session,  $F(1,21) = 7.61, p = 0.01$ , shown in Fig. 6. Planned  $t$ -tests revealed that the trained group improved on the reading span measure,  $t(11) = 3.46, p = 0.005$  (pre-test  $M = 6.33$  words, post-test  $M = 7.42$  words), but the control group did not,  $t(10) = 0.32, p = 0.76$  (pre-test  $M = 6.09$  words, post-test  $M = 6.00$  words). The main effects of group  $F(1,21) = 9.44, p = 0.005$ , and session,  $F(1,21) = 6.02, p = 0.02$ , were also significant, but these main effects should be interpreted with an eye to the significant interaction.

To compare the results from our English-speaking listeners to those from our Mandarin-speaking listeners, we converted the QuickSIN results from SNR loss to SRT-50 (Killian *et al.*, 2004). Because the M-HINT starts at  $-4$  dB

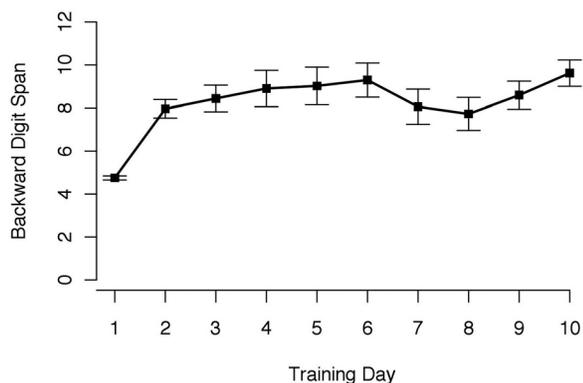


FIG. 4. Mean backward digit span performance for the native English speakers as a function of training day. Training conditions were the same as in the Mandarin experiment. Error bars are standard error of the mean.

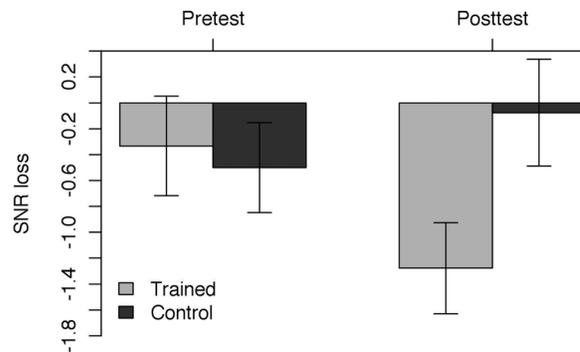


FIG. 5. Mean SNR loss on the QuickSIN at pre- and post-test by the trained and control native English speaking subjects. Trained subjects showed a significant improvement on their ability to perceive speech in noise, whereas control subjects did not. Error bars are standard error of the mean.

SNR then adapts the presentation level based on the listener's performance, whereas the QuickSIN starts at 25 dB SNR then changes the SNR level in 5 dB decrements, comparisons across language groups were made using gain scores, shown in Fig. 7. For trained groups, no significant differences were found between Mandarin- and English-speaking listeners' gains in speech perception in noise,  $t(25) = 0.222, p = 0.83$  (Mandarin RTS  $M = -0.84$  dB SNR; English SRT-50  $M = -0.75$  dB SNR), or reading span,  $t(25) = 0.710, p = 0.48$  (Mandarin  $M = 1.47$  words; English  $M = 1.08$  words). The control groups also showed no difference on gains in speech perception in noise,  $t(23) = 1.71, p = 0.10$  (Mandarin RTS  $M = -0.28$  dB SNR; English SRT-50  $M = 0.42$  dB SNR), nor on reading span gains,  $t(23) = 0.728, p = 0.047$  (Mandarin  $M = 0.14$  words; English  $M = -0.09$  words).

## C. Discussion

As with the Mandarin speakers, we again saw significant gains on training, reading span, and speech perception in noise performance. In addition, this second experiment supports our hypothesis that using digits as training materials is easily transferable among languages as speech perception gains of similar magnitudes were seen in both Mandarin and English listeners.

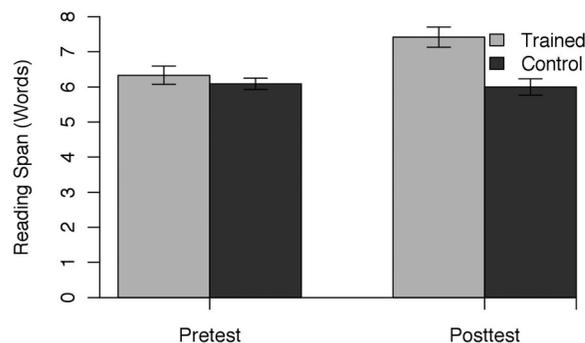


FIG. 6. Mean reading span performance by the native English speaking subjects. Trained subjects significantly increased their reading spans following training, whereas there was no change in control spans at posttest. Error bars are standard error of the mean.

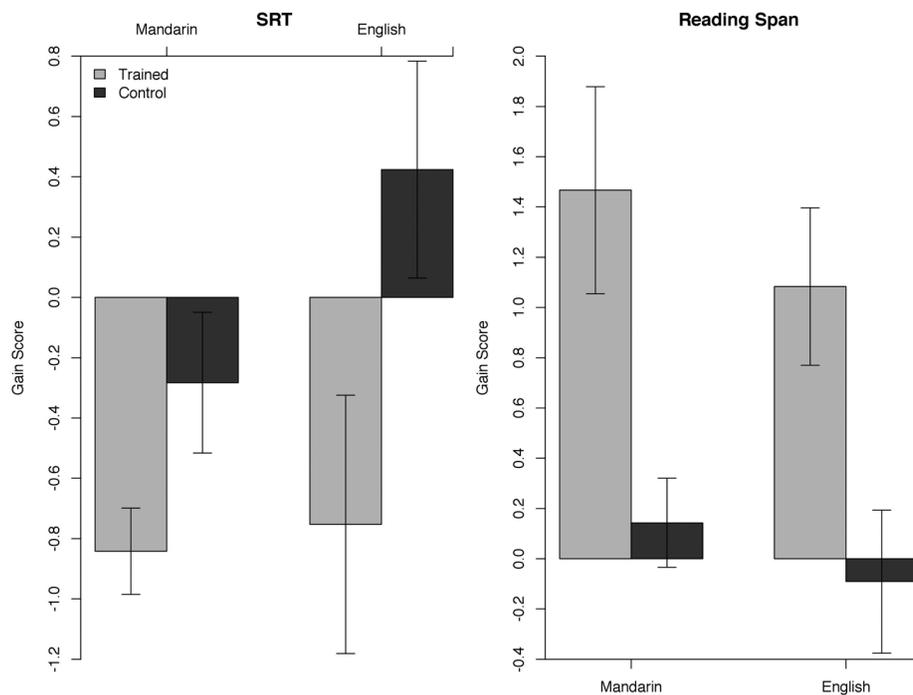


FIG. 7. Gain scores for both the Mandarin- and English-speaking subjects for both the speech-in-noise and reading-span tasks. To compare gain scores across groups, English-speakers' QUICKSIN performance, measured in SNR loss, was converted to SRT. There were no significant differences in performance gain for trained or control listeners on either task. Error bars are standard error of the mean.

#### IV. GENERAL DISCUSSION

Difficulties perceiving speech in noisy situations are a relatively common complaint, exhibited by older adults with normal hearing, hearing aid users, and users of cochlear implants. Faced with growing numbers of older adults and limitations in hearing technologies, research has turned to rehabilitative possibilities to improve speech perception in noise. Noting that listeners with greater working memory capacity showed better speech perception in noise (Foo *et al.*, 2007; Lunner and Sundewall-Thorén, 2007), we hypothesized that training working memory capacity would be associated with improved speech perception in noise.

In addition to our hypothesis that training working memory would be associated with improved speech perception in noise, we hypothesized that using digits as the training stimuli would be easily transferable across languages. Currently commercially available speech perception in noise training programs are based in linguistic materials, making it a very labor-intensive endeavor to translate these programs into languages other than English. We hypothesized that digits would allow us to see post-treatment gains in multiple languages, potentially leading to interventions in a variety of languages, allowing listeners who are not proficient in English to improve their speech perception in noise.

Both of our hypotheses were confirmed. Working memory training was associated with improved speech perception in noise relative to untrained controls for both native Mandarin and native English speakers. In both cases, gains were made by young, normal hearing, cognitively healthy listeners. Though we expect our training regime to be effective for improving speech perception in noise for older adults and listeners with hearing loss, we recognize that this expectation is pure speculation and remains to be explicitly tested. We base this expectation on earlier findings that older adults may receive a larger benefit from cognitive training than

younger adults (Bherer *et al.*, 2005; Li *et al.*, 2008). Although there has been some suggestion that young, healthy participants show larger training effects than older participants (Dahlin *et al.*, 2008; Schmiedek *et al.*, 2010), when older adults are assessed using measures that are ecologically valid, they show transfer effects that are on par with those of younger adults (Morrison and Chein, 2010). Because speech perception in noise is an ecologically valid task for older adults and listeners with hearing loss (Gordon-Salant and Fitzgibbons, 1997) and that other studies have found speech-perception-in-noise gains in listeners with hearing loss (Henshaw and Ferguson, 2013; Ingvalson *et al.*, 2013b; Miller *et al.*, 2008), we are confident that both older adults and listeners with hearing loss will show speech-perception-in-noise gains following training. We anticipate that certain modifications will need to be made to the training paradigm to make it appropriate for older adults and listeners with hearing loss; we discuss these modifications in more detail in the following text.

A major limitation of the current study was the control group. Because the control group was no-contact, they spent 83% less time in the lab than the trained listeners. We cannot therefore say with certainty that gains made by the trained listeners are solely a result of training and not a result of more familiarity with the testing apparatus, more practice with the experimental environment, or an awareness of experimenter expectation (McCarney *et al.*, 2007). More crucially, given that the training paradigm contained a speech-perception-in-noise component, we cannot say how much of the speech-perception-in-noise gain the trained listeners made was a result of the working memory gains or from practice listening to speech in noise. A stronger design would be to compare the current training paradigm to a relatively easy, nonadaptive digit span paired with the speech-in-noise regimen as well as to an adaptive digit span without

noise to determine the contribution practice perceiving speech in noise vs increasing working memory capacity contributed to the present gains and whether the combined approach used here would result in optimal performance.

In the present study, we paired an adaptive working memory training with a set SNR level and noise type; this study was done with a set of fairly homogeneous group of university students with normal hearing and relatively high working memory capacities. We regularly changed the noise type and SNR ratio to maintain task difficulty. Our strategy of maintaining task difficulty is consistent with the literature demonstrating improved learning when learners are challenged (Amitay *et al.*, 2006; Klingberg, 2010). However, this same literature has demonstrated that there is an optimum level of difficulty: if learners are too challenged or not challenged enough, then learning will not take place (Morrison and Chein, 2010). In the current study, we began to approach the limits of difficulty for effective learning, evidenced in the  $-10$  dB SNR speech-shaped-noise condition for the English listeners where we saw a slight drop in performance. In contrast with the current study, investigations with older adults are likely to use a population that is more heterogeneous with regard to hearing ability, speech-perception-in-noise ability or working memory capacity (Koelewijn *et al.*, 2014). This difference in population heterogeneity is important when considering the future of our intervention. An intervention that has its difficulty set somewhat arbitrarily by the experimenter or clinician is appropriate for a homogeneous population as was used here, but it is not likely to be so for the heterogeneous population that will seek treatment for speech perception in noise difficulties. Future interventions will likely have broader appeal if the working memory difficulty and the SNR difficulty both adapt based on the performance of the listener. The two difficulty levels should adapt independently, reflecting both the listener's ability to do the working memory task and the listener's ability to perceive speech in noise (Henshaw and Ferguson, 2013; Ingvalson *et al.*, 2013b).

Although relatively small, 0.80 dB SNR averaged for the two language groups, the magnitude of the gains seen following our training are comparable to gains following other speech-perception-in-noise trainings (Miller *et al.*, 2008; Sweetow and Sabes, 2006). The small magnitude of these, and previous, results, raises the question of the clinical benefit of speech-perception-in-noise training. To address this question, we turn to the gain on our measure of working memory where listeners showed a gain of 1.30 words averages across the language groups. Lunner and Sundewall-Thorén found that working memory capacity accounted for 39% of the variance in speech perception in noise among hearing aid users but degree of hearing loss only accounted for 3% (2007). We therefore argue that the speech-perception-in-noise gains from our current training are likely to be clinically significant because they stem from working memory gains and thereby represent an increase in cognitive resources available for speech perception rather than an improvement in hearing acuity (Rönnerberg *et al.*, 2008).

One under-explored avenue for determining clinical benefit is through listener self-report (Henshaw and

Ferguson, 2013). As we begin to investigate the potential benefit of our intervention in older adults and listeners with hearing loss, bearing in mind the modifications discussed in the preceding text, we will want to determine the extent to which listeners perceive a benefit following training. Subjective measures will provide an indication of the relationship between decibel RST or decibel SRN loss gains and the gain that listeners perceive in their daily life. Similarly, just as other researchers have found relationships between working memory capacity, listening effort, and speech perception in noise (Koelewijn *et al.*, 2014; Zekveld *et al.*, 2011), it may be possible to relate changes in working memory capacity to changes in listening effort, possibly through pupillometry. Using subjective measures of benefit and objective measures of listening effort will help us to determine the extent to which listeners are receiving a qualitative gain from training. Understanding the extent to which listeners are receiving a qualitative gain from training will allow us to test our hypothesis that the seemingly small speech-perception-in-noise gains from the current intervention will prove to be clinically meaningful because they stem from increases in working memory capacity. Additionally considering that many listeners report a lack of speech perception in noise benefit from hearing aids (Killion, 2004), ensuring that gains are clinically meaningful in addition to statistically significant will be an important step for the future.

In addition to the clinical benefit, how long gains can be maintained, or what the optimal dosage is for training, is also unclear. The training dosage used in the current study was short (10 days) and was consistent with training durations we used in previous experiments (Ingvalson *et al.*, 2012; Ingvalson *et al.*, 2013a). Studies investigating the effectiveness of commercial speech perception in noise training have used much longer training durations, typically around 4 wk (Miller *et al.*, 2008; Sweetow and Sabes, 2006). Although the durations are very different, the commonality among the studies is that the duration of training was set by the experimenter. In the current study, participants' spans continued to climb throughout training (Figs. 1 and 4). In the interests of optimizing individual listeners' performance, one approach might be to have listeners train to asymptotic performance rather than for a set number of days. Further work is needed to determine at which point to terminate training to maximize outcomes, but it may be several days of consistent spans at an unchanging SNR level. Assuming the intervention moves to a model where the SNR level is set by the listeners' performance, it may be that asymptotic performance is set by several days of consistent spans within a relatively narrow SNR range.

For both long and short training durations, whether gains are maintained after training finishes is unknown. With increasing interest in cognitive training, it has been shown that gains can be maintained for up to 6 wk post-intervention (Klingberg, 2010). Beyond that, trainees may require occasional "tune-ups" to keep their skills current (Ball *et al.*, 2002). How long gains can be maintained following the present intervention—for both younger and older listeners with and without hearing loss—as well as what a maintenance intervention might look like will need to be further explored.

Here we demonstrated that simple working memory training was associated with significantly improved speech perception in noise and that the training stimuli could be easily adapted across languages, and evidence of improvement was seen for both native Mandarin and native English listeners. Although there remains much work to be done before these data can be used to aid clinical practice, we are optimistic that this future work will prove fruitful and that working memory interventions will result in better speech perception in noisy situations for all listeners.

## ACKNOWLEDGMENTS

The authors wish to thank Weifeng Li and Casandra Nowicki for their help with data collection. Our research is supported by grants from the National Institutes of Health (R01DC008333 to P.C.M.W.), the Research Grants Council of Hong Kong (GRF477513 to P.C.M.W.), the Stanley Ho Medical Development Foundation (to P.C.M.W.), University Grants Committee (HKSAR) (477513, 14117514 to P.C.M.W.), the Global Parent Child Resource Centre Limited, the National Nature Science Foundation of China (31371135 to H.L.), and Guangdong Natural Science Funds for Distinguished Young Scholar (S2013050014470 to H.L.).

- Amitay, S., Irwin, A., and Moore, D. R. (2006). "Discrimination learning induced by training with identical stimuli," *Nat. Neurosci.* **9**, 1446–1448.
- Baddeley, A. (2003). "Working memory: Looking back and looking forward," *Nat. Rev. Neurosci.* **4**, 829–839.
- Baddeley, A. D. (1986). *Working Memory* (Oxford University Press, Oxford, UK), 230 pp.
- Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., and Willis, S. L. (2002). "Effects of cognitive training interventions with older adults: A randomized controlled trial," *J. Am. Med. Assoc.* **288**, 2271–2281.
- Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., and Becic, E. (2005). "Training effects on dual-task performance: Are there age-related differences in plasticity of attentional control?," *Psychol. Aging* **20**, 695–709.
- Brouwer, S., Van Engen, K. J., Calandruccio, L., and Bradlow, A. R. (2012). "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content," *J. Acoust. Soc. Am.* **131**, 1449–1464.
- Cabeza, R., Dolcos, F., Graham, R., and Nyberg, L. (2002). "Similarities and differences in the neural correlates of episodic memory retrieval and working memory," *Neuroimage* **16**, 317–330.
- Chen, J. M., and Morrison, A. B. (2010). "Expanding the mind's workspace: Training and transfer effects with a complex working memory span task," *Psychon. Bull. Rev.* **17**, 193–199.
- Dahlin, E., Nyberg, L., Bäckman, L., and Neely, A. S. (2008). "Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance," *Psychol. Aging* **23**, 720–730.
- Daneman, M., and Carpenter, P. A. (1980). "Individual differences in working memory and reading," *J. Verbal Learn. Verbal Behav.* **19**, 450–466.
- Daneman, M., and Merikle, P. M. (1996). "Working memory and language comprehension: A meta-analysis," *Psychon. Bull. Rev.* **3**, 422–433.
- Duncan, K. R., and Aarts, N. L. (2006). "A comparison of the HINT and Quick Sin Tests," *J. Speech Lang. Pathol. Audiol.* **30**, 86.
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state. A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatric Res.* **12**(3), 189–198.
- Foo, C., Rudner, M., Rönnerberg, J., and Lunner, T. (2007). "Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity," *J. Am. Acad. Audiol.* **18**, 618–631.
- Garcia Lecumberri, M. L., and Cooke, M. (2006). "Effect of masker type on native and non-native consonant perception in noise," *J. Acoust. Soc. Am.* **119**, 2445–2455.
- Gordon-Salant, S., and Fitzgibbons, P. J. (1997). "Selected cognitive factors and speech recognition performance among young and elderly listeners," *J. Speech Lang. Hear. Res.* **40**, 423–431.
- Henshaw, H., and Ferguson, M. A. (2013). "Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence," *PLoS ONE* **8**, e62836.
- Hickok, G., and Poeppel, D. (2007). "The cortical organization of speech processing," *Nat. Rev. Neurosci.* **8**, 393–402.
- Ingvalson, E. M., Barr, A. M., and Wong, P. C. M. (2013a). "Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning," *J. Speech Lang. Hear. Res.* **56**, 1045–1050.
- Ingvalson, E. M., Holt, L. L., and McClelland, J. L. (2012). "Can native Japanese listeners learn to differentiate /r-/l/ on the basis of F3 onset frequency?," *Biling. Lang. Cogn.* **15**, 255–274.
- Ingvalson, E. M., Lee, B., Fiebig, P., and Wong, P. C. M. (2013b). "The effects of short-term computerized speech-in-noise training on post-lingually deafened adult cochlear implant recipients," *J. Speech Lang. Hear. Res.* **56**, 81–88.
- Ingvalson, E. M., and Wong, P. C. M. (2013). "Training to improve language outcomes in cochlear implant recipients," *Front. Psychol.* **4**, 263.
- Killion, M. C. (2004). "Myths about hearing aid benefit and satisfaction," *Hear. Rev.* **11**, 14–21.
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **116**, 2395–2405.
- Kinsella, K., and He, W. (2008). "An Aging World: 2008," International Population Reports, U.S. Census Bureau.
- Klingberg, T. (2010). "Training and plasticity of working memory," *Trends Cogn. Sci.* **14**, 317–324.
- Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2014). "The influence of informational masking on speech perception and pupil response in adults with hearing impairment," *J. Acoust. Soc. Am.* **135**, 1596–1606.
- Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., and Lindenberger, U. (2008). "Working memory plasticity in old age: Practice gain, transfer, and maintenance," *Psychol. Aging* **23**, 731–742.
- Lunner, T. (2003). "Cognitive function in relation to hearing aid use," *Int. J. Audiol.* **42**, S49–S58.
- Lunner, T., and Sundewall-Thorén, E. (2007). "Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid," *J. Am. Acad. Audiol.* **18**, 604–617.
- McArdle, R. A., and Wilson, R. H. (2006). "Homogeneity of the 18 QuickSIN lists," *J. Am. Acad. Audiol.* **17**, 157–167.
- McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., and Fisher, P. (2007). "The Hawthorne Effect: A randomised, controlled trial," *BMC Med. Res. Method.* **7**, 30.
- Melby-Lervåg, M., and Hulme, C. (2013). "Is working memory training effective? A meta-analytic review," *Dev. Psychol.* **49**, 270–291.
- Miller, J. D., Watson, C. S., Kistler, D. J., Wightman, F. L., and Preminger, J. E. (2008). "Preliminary evaluation of the speech perception assessment and training system (SPATS) with hearing-aid and cochlear-implant users," *Proc. Meet. Acoust.* **2**, 1–9.
- Morrison, A. B., and Chen, J. M. (2010). "Does working memory training work? The promise and challenges of enhancing cognition by training working memory," *Psychon. Bull. Rev.* **18**, 46–60.
- O'Bryant, S. E., Waring, S. C., Cullum, C. M., Hall, J., Lacritz, L., Massman, P. J., Lupo, P. J., Reisch, J. S., and Doody, R. (2008). "Staging dementia using clinical dementia rating scale sum of boxes scores," *Arch. Neurol.* **65**(8), 1091–1095.
- Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). "Musician enhancement for speech-in-noise," *Ear Hear.* **30**, 653–661.
- Persson, J., and Reuter-Lorenz, P. A. (2008). "Gaining control: Training executive function and far transfer of the ability to resolve interference," *Psychol. Sci.* **19**, 881–888.
- Rönnerberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). "Cognition counts: A working memory system for ease of language understanding (ELU)," *Int. J. Audiol.* **47**, S99–S105.

- Schmiedek, F., Lovden, M., and Lindenberger, U. (2010). "Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study," *Front. Aging Neurosci.* **2**, 27
- Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). "Is working memory training effective?," *Psychol. Bull.* **138**, 628–654.
- Smits, C., Goverts, S. T., and Festen, J. M. (2013). "The digits-in-noise test: Assessing auditory speech recognition abilities in noise," *J. Acoust. Soc. Am.* **133**, 1693–1706.
- Sweetow, R. W., and Sabes, J. H. (2006). "The need for development of an adaptive listening and communication enhancement (LACE) program," *J. Am. Acad. Audiol.* **17**, 538–558.
- Van Engen, K. J., and Bradlow, A. R. (2007). "Sentence recognition in native- and foreign-language multi-talker background noise," *J. Acoust. Soc. Am.* **121**, 519–526.
- Weber, J., Mueller, H. G., and Johnson, E. (2010). *Fitting Hearing Aids: A Comparison of Three Pre-Fitting Speech Tests*, AudiologyOnline, Available: <http://www.audiologyonline.com/articles/fitting-hearing-aids-comparison-three-861> (Last viewed 2/13/2015).
- Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., Morris, J. N., Rebok, G. W., Unverzagt, F. W., Stoddard, A. M., and Wright, E. (2006). "Long-term effects of cognitive training on everyday functional outcomes in older adults," *JAMA J. Am. Med. Assoc.* **296**, 2805–2814.
- Wilson, R. H., McArdle, R., Betancourt, M. B., Herring, K., Lipton, T., and Chisolm, T. H. (2010). "Word-recognition performance in interrupted noise by young listeners with normal hearing and older listeners with hearing loss," *J. Am. Acad. Audiol.* **21**, 90–109.
- Wilson, R. H., McArdle, R. A., and Smith, S. L. (2007). "An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss," *J. Speech Lang. Hear. Res.* **50**, 844–856.
- Wong, L. L. N., Liu, S., and Han, N. (2008). "The mainland Mandarin hearing in noise test," *Int. J. Audiol.* **47**, 393–395.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). "Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response," *Ear Hear.* **32**, 498–510.