Auditory Training for Adults Who Have Hearing Loss: A Comparison of Spaced Versus Massed Practice Schedules

Nancy Tye-Murray, PhD[1]

Brent Spehar, PhD[1]

Joe Barcroft, PhD[2]

Mitchell Sommers, PhD[2]

[1]Washington University in St Louis

[2]Washington University in St Louis School of Medicine

Author Note

Support was provided by grant RO1DC008964 from the National Institutes of Health.

Correspondence should be directed to Nancy Tye-Murray, Washington University in St. Louis School of Medicine, MO Department of Otolaryngology 63110; email: murrayn@wustl.edu

ABSTRACT

Purpose: The *spacing effect* in human memory research refers to situations in which people learn items better when they study items in spaced intervals as compared to studying items in massed intervals. This investigation compared the efficacy of meaning-oriented auditory training when administered with a spaced versus massed practice schedule.

Methods: Forty-seven adult hearing aid users received 16 hours of auditory training. Participants in a spaced group (mean age = 64.6, SD = 14.7) trained twice a week whereas participants in a massed group (mean age = 69.6, SD = 17.5) trained five consecutive days each week. Participants completed speech perception tests before training, immediately following training, and then 3 months later. Consonant with transfer appropriate processing theory, tests assessed both trained tasks and an untrained task. included 10 females.

Results: Auditory training improved the speech recognition performance of participants in both groups. Benefits were maintained for 3 months. There was no effect of practice schedule on overall benefits achieved, on retention of benefits, nor on generalizability of benefits to non-trained tasks.

Conclusions: The lack of spacing effect in otherwise effective auditory training suggests that perceptual learning may be subject to different influences than are other types of learning, such as vocabulary learning. Hence, clinicians might have latitude in recommending training schedules so as to accommodate patients' schedules.

INTRODUCTION

Auditory training attempts to teach individuals with hearing loss to maximally utilize their residual hearing for recognizing speech. Participants are presented with a series of learning exercises and benefits are assessed by comparing performance prior to and then following the training protocols. Recently, auditory training programs have moved from a focus on discrimination of phonological form (e.g., identify the difference between syllables such as /ba/ and /pa/) to ones that emphasize demands required for real-world communication. For example, Humes et al. (2009) incorporated word frequency metrics into the design of their training program and restricted training to the 600 most frequent words in English that constitute approximately 90% of all words encountered during typical conversations. Barcroft et al. (2011) incorporated meaning-oriented training into their auditory training program, which requires clients to map linguistic forms to their meaning (e.g., distinguish between word pairs such as *bat* and *pat* by identifying pictures to which they refer). In the present investigation, we examine how one of the most extensively investigated learning principles, the *spacing effect*, applies to meaning-oriented auditory training.

The *spacing effect* in research on human learning and memory refers to how, in some instances, we learn items better when learning episodes are distributed across a given time interval as compared to when they are presented as a massed group. First documented by Ebbinghaus (1885), who pointed out that more distributed repetitions lead to increased learning/memory, the robustness of the spacing effect has been confirmed in a variety of different learning contexts (see Green & Bavelier, 2008, for a review). In one classic study, for example, Melton (1967) demonstrated that single words presented in lists were remembered better if their occurrences were separated by other words than if they occurred consecutively. Bahrick, Bahrick, Bahrick, and Bahrick (1993) demonstrated that more spaced practice (as compared to more massed practice) resulted in better second-language vocabulary-learning and maintenance over a 5-year period. The spacing effect has also been seen in long-term recall tasks, such as verbal learning (Bahrick et al., 1993), conditioning (Carew, Pinkster, & Kandel, 1972), and learning of academic material (Dempster, 1988). In addition, a spacing effect has been documented in studies of cognitive training with children (e.g., Wang, Zhou, & Shah, 2014).

Spacing not only affects learning of material but also affects the extent to which learning generalizes to other contexts. For example, using a computerized “Space Fortress” game, Shebilske, Goettl, Corrington, and Day (1999) found that spaced training led to better skill acquisition than massed training. Participants in the spaced-training group also demonstrated better transfer of skill from a joystick to a keyboard, suggesting that spaced training leads to better generalizability. Additionally, based on a review on spacing effects on categorization and generalization of learning among children, Vlach (2014) explains how periodic forgetting may play a role in the emergence of benefits of spacing on generalization of learning.

The *distributed practice effect* refers to the influence of both spacing and lag effects (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The *lag effect* is the effect that the duration of an inter-study interval (ISI) has upon learning. A series of meta-analyses suggest that longer ISIs enhance learning of verbal information (such as spelling and foreign language learning), working memory, and retention of motor skills (such as typing and gymnastics) (Moss, 1996; Lee & Genovese, 1988; Cepeda et al., 2006; Wang, et al., 2014).

Little research has addressed the possibility that the efficacy of auditory training is affected by a spacing effect even though published reports about auditory training present widely-varying distribution schedules of practice. For example, training may occur daily (e.g., Stacey et al., 2010), five days a week (e.g., Fu, Galvin, Wang, & Nogaki 2004; Stecker et al., 2006; Sweetow & Henderson-Sabes, 2006; Oba, Fu, & Galvin, 2011; Zhang, Dorman, Fu, & Spahr, 2012), three days a week (Burk & Humes, 2008), or twice a week (Barcroft et al., 2011; Barcroft, Spehar, Tye-Murray, & Sommers, 2016; Tye-Murray, Spehar, Sommers, & Barcroft, 2016). Published reports also vary widely in reporting the extent to which auditory training improves the speech recognition performance of adults who have hearing loss (e.g., see Tye-Murray, 2015, for a review; see also Henshaw & Ferguson 2013, and Sweetow & Palmer, 2005, for meta-analyses showing a paucity of evidence-based support for efficacy). One factor that might contribute to the varying findings is that researchers fail to take into account the effect of practice schedules. That is, some schedules might promote auditory learning whereas others may not.

Interestingly, extant research on the effects of massed versus spaced training on speech perception and other areas of perceptual learning is very limited. Although previous studies have demonstrated spacing effects for a wide variety of types of learning, including for learning novel shapes (Cornoldi & Longoni, 1977) and novel faces (Russo, Parkin, Taylor, & Wilkes, 1998), little research has focused on whether the spacing effect generalizes to contexts in which individuals attempt to improve their perceptual acuity in domains such as vision and audition. Given evidence for the differential effects of massed versus spaced training for different types of memory, including object-identity and object-location recognition memory in non-human animals (Bello-Medina, Sánchez-Carrasco, González-Ornelas, Jeffrey, & Ramirez-Amaya, 2013), a certain degree of skepticism is warranted when it comes to the hypothesis that spaced training should be superior to massed training for perceptual learning. Some evidence suggests that increased time for sleep promotes visual discrimination performance, one area of perceptual learning, but different amounts of sleep is not the same as a massed versus spaced training in and of itself. Given this larger picture and the very limited amount of previous research on massed versus spaced perceptual learning, the present study on the potential effects of massed versus spaced training on auditory training makes for a unique test case regarding the viability (or lack thereof) of the spacing effect in the realm of perceptual learning.

At least one study has considered the issue of whether spacing and distributed practice might affect the outcome of auditory training. In a study that entailed auditory perceptual learning, Molloy, Moore, Sohoglu, and Amitay (2012) used a frequency discrimination task to compare single-session training to multi-day training. Their participants were trained with computer games to discriminate tones varying in frequency, using a three-interval, forced-choice paradigm. Their findings were confounded by the fact that those participants enrolled in the multi-day training group received a longer period of training and more exposure to the training tasks. They suggest that their findings indicate that training should be presented in shorter sessions that are optimally spaced over time. They also noted that participants who received training in a single session continued to improve, even after training ceased, whereas participants in a multi-session group did not continue to improve.

There are at least three ways that a distributed practice schedule might affect the outcome of an auditory training program. The first and most obvious effect is on overall benefit. Students who receive spaced training might demonstrate the classic spacing effect and attain more benefit than students who receive massed training. Arguably, however, perceptual learning—in the case of auditory training, learning to recognize a degraded auditory signal—is qualitatively different than learning a new motor task or new verbal or academic material, especially by learners who are native speakers of the language being trained. It is therefore possible that when learning to recognize degraded auditory speech signals, closely spaced training might be most effective because the learner has a better opportunity to remember, compare, contrast, and ultimately recognize auditory patterns associated with spoken vocabulary. If this is indeed the case, individuals with hearing loss who receive massed auditory training might realize greater gains in speech recognition than those who receive spaced training. In this investigation, participants were enrolled in either a spaced or massed training schedule, allowing us to assess whether spaced or massed auditory training would result in more overall benefit. In light of demonstrations of the spacing effect across a wide range of learning contexts, our hypothesis was that spaced training would result in greater overall benefit when compared to massed training. Note also that if such a benefit were to be observed (or not), it likely would not be related to the older ages of the participants in our study (as compared to younger adults) considering the large amount of previous research showing benefits for spacing over massed training in both younger and older adult populations (see, e.g., Kornell et al., 2010, for an example and review) although the magnitude of the benefit may diminish among older adults (Simone, Bell & Cepeda, 2012).

The second way that a practice schedule might affect the efficacy of auditory training concerns retention. It may be that as with other perceptual and motoric tasks, spaced practice leads to better retention of benefits (e.g., Cepeda et al., 2006). The present study therefore included delayed measures, allowing us to assess whether spaced training might lead to greater retention when compared to massed training. Although speculative given lack of previous research in this area, we hypothesized that spaced training might lead to greater retention of the benefits of auditory training.

Finally, the third way that a practice schedule might affect the efficacy of auditory training concerns generalizability. In previous work, we adopted a general transfer-appropriate processing (TAP) perspective with regard to understanding the generalizability of benefits of auditory training wherein we posited that increased similarity in task, talker, and stimuli from training to assessment would produce increased gains (Barcroft et al., 2011; Barcroft et al., 2016). For example, Barcroft et al. (2011; see also Barcroft et al., 2016) found that participants trained in a single-talker condition improved significantly more on a four-choice discrimination task when their post-training test stimuli were spoken by the same talker with whom they had trained than when spoken by a different talker or by multiple talkers. Such findings speak to the importance of degree of overlap between the component perceptual and cognitive processes engaged during training and what can be reasonably expected when it comes to "generalizability" per se: the greater the overlap, the greater the "generalizability" of benefit. Indeed, our most current work with auditory training applies this processing-specific approach by providing talker-specific auditory training. Tye-Murray et al. (2016) demonstrated that adults who use hearing aids improved their discrimination of speech produced by their spouses after having received auditory training with stimuli spoken by their spouse, despite having been married an average of 14 years. Might massed increase generalizability by enhancing processing-specific gains of this nature? While speculative, our position was that if spaced training led to greater gains overall, then it would also lead to more generalizability when viewed from this perspective of degree of overlap among processes engaged at study and at test.

In this investigation, we compared how two distributed practice schedules affect overall benefit of auditory training, retention of benefits, and generalization of training benefits. We considered two practice schedules that have been used in previous investigations: the five-days-of-the-week schedule to complete 16 hours of training (e.g., Fu et al., 2004; Stecker et. al, 2006; Sweetow & Henderson-Sabes, 2006; Oba et al., 2011; Zhang et al., 2012) and the twice-weekly schedule (e.g. Barcroft et al., 2011; Barcroft et al., 2016), called *massed training* and *spaced training* in this report, respectively. We controlled for the amount of training received by participants in each training group, both in terms of time-on-task and in terms of exposure to number of training exercises. Moreover, we also equated the time between the final training period and the post-test assessment. As suggested by the review above regarding potential differences between massed and distributed training, we investigated whether the different training schedules would affect (1) initial learning, (2) extent of generalization, and (3) 3-mos retention. Participants were tested before, immediately after completing 16 hours of auditory training, and then again 3 months following the end of training. Although it is certainly possible that massed versus spaced training would have differential effects on our outcome measures (learning, generalization, retention), we predicted improved performance on all three for spaced compared to massed training.

METHODS

Participants

To be eligible for the study, participants had to report not having received formal auditory training for six months prior to enrollment. Fifty-nine adult hearing aid users with sensorineural hearing loss agreed to take part in the study. Forty-seven participants completed the entire experimental protocol and their data are reported here. Eleven participants withdrew from the study following their Pre training assessment session after learning about the time commitment involved and about the restrictions that participating in auditory training would place upon their daily schedules. One participant did not return after the Post training assessment.

To create approximately equal group sizes, we assigned participants to two training groups (massed or spaced) in a counterbalanced manner whereby every other volunteer was assigned to the spaced training group. The 24 participants in the spaced training group (mean age = 64.6, SD = 14.7) included 10 females. The 23 participants in the massed training group (mean age = 69.6, SD = 17.5) included 12 females. The age difference between the two groups was not statistically significant (t (45) = 1.3, p = .195). Hearing thresholds were measured by taking the average unaided threshold in dB HL for pure tones presented at 500, 1000 and 2000 Hz (PTA). The mean PTAs for each participant's better ear was not statistically significantly different between the two groups (mean PTA for spaced = 41.4 dB HL, SD = 14.7; mean PTA for massed = 42.7 dB HL, SD = 17.5) (t (45) = .28, p = .774). We recruited all participants through the Volunteers for Health (VFH) program at Washington University School of Medicine and received $10/hour for their participation. VFH is a service maintained by the university hospital to connect, through advertisement, potential research participants with studies that may be of interest to them. We screened potential participants via questionnaire to exclude those who had had neurological events such as stroke, open or closed head injury. In order to screen for dementia, participants completed the Mini Mental Status Exam (MMSE, Folstein, Folstein, & McHugh, 1975).

Training

The training schedule is presented in Figure 1. All auditory training was provided in sound treated booths located in the speech and hearing laboratories at Washington University school of Medicine in St Louis. All participants completed 20 one-hour sessions. Each participant received the same sequence of training activities and stimuli throughout the 20 sessions. Participants in the massed training group returned five times each week for two weeks and completed two sessions at each visit with a rest break provided between the two sessions. Participants in the spaced training group returned twice each week for ten weeks and completed only one session per visit. During training, the participant sat a comfortable distance from a 19” ELO Touchsystems monitor and heard training stimuli presented through two loudspeakers positioned at $\pm 45^{o}$.

---Insert Figure 1 Here---

Each session included five training activities taken from the catalogue of activities from customized learning: Exercises for Aural Rehabilitation (clEAR, formerly *I Hear What You Mean*; see, e.g., Tye-Murray, Sommers, & Barcroft, 2011). Four of the games used adaptive presentations of the target’s audio while in the presence of six-talker babble to maintain a constant level of difficulty. The background babble was always presented at approximately 62 dB SPL. The target audio for the fifth game, the Build-a-Paragraph game (discussed below), was presented in quiet. The level of the target speech was adjusted after each trial based on the listener’s responses using a two-down one-up procedure to keep performance at approximately 71% on the first response to each trial (Levitt, 1970). For example, when participants selected the wrong answer for the first attempt of a trial, the item would disappear from the screen and the trial would be repeated at a level that was at a signal-to-babble ratio (SBR) that was 2dB easier until the correct answer was selected. If a correct response was given on the first attempt for two trials in a row, the next presentation was at a SBR that was 2dB harder. Three male and three female actors participated as talkers for the training material. Talkers were rotated after every trial during each activity. No trial or set of stimuli was repeated throughout the 20 one-hour sessions.

Each training game/activity took approximately 10-12 minutes to complete. Activity 1 was a four-choice *word identification* task that required participants to choose among four pictures on the screen that best represents the word that was presented (i.e., *zoom*, *room*, *tomb*, *cat*). Activity 2, a *four-choice discrimination* task, also required participants to choose among four pictures on the screen. However, the presentations were matched-pairs of words and the participant chose which pair represented the auditory presentation (i.e., *buy-pie*, *buy-buy*, *pie-buy*, or *pie-pie*). Activity 3 was a *fill-in-the-blank* task where participants listened to a sentence that was missing the final word and then choose, among four auditory alternatives, the word that best completed the sentence (i.e., *The grass had gotten long so he had to [‘mow’, ‘know’, ‘row’, ‘low’*].). Activity 4 was a three-choice *sentence context* task that asked participants to choose among three sentences provided via text on the screen that would best compliment the sentence that was presented auditorily. For example, when the auditory presentation was, *Bob asked the waiter to remove the bug from the table,* the participant chose from the following options that appeared orthographically on the screen: *He had finished his coffee* (which might be selected if the participant had heard the word *mug* instead of *bug*), *The people sitting nearby had brought their puppy* (which might be selected if the participant had heard the word *pug*), and *It was a firefly* (correct answer). Finally, Activity 5 was a comprehension task that required participants to first listen to paragraphs that included five-sentences and then rearrange the same sentences until they were in the same order that was originally presented. Participants listened to each sentence one at a time and chose the order number for that sentence. Paragraphs were constructed so as to only have one logical sequence (e.g., *Bill and Sandy planned a picnic. Sandy made the shopping list. Sandy gave the list to Bill. Bill drove to the store. Bill bought the items on the list*.). Participants were allowed to listen to each sentence and reorder their answer as many times as needed before making a final decision.

Outcome Measures

Outcome measures were administered at three assessment times. A pre-training session (Pre) was conducted before training began, a post-training assessment (Post) was administered at the completion of training, and a three-months-post training (3 Mos) assessment was administered three months after the Post assessment. The spaced training group began training within a week of completing the Pre assessment whereas the massed training group waited 8 weeks before beginning training in order to ensure that the Post assessment was competed at the same time interval for both groups.

The outcome measures represent two distinct types of assessments. First are items specifically chosen to represent a TAP-style measure of improvement (Barcroft, Sommers, Tye-Murray et. al., 2011; Barcroft et. al., 2016). The second type of assessment is to measure the generalizability of training to other tasks and stimuli.

The TAP assessment package was comprised of four tasks that were similar to the tasks used during training (i.e.,, word identification, four-choice discrimination, fill-in-the-blank, and sentence context tasks). An even distribution of the same talkers used in training was also used in the assessments. For the assessment, the level of the target speech was held constant at 60 dB SPL and the six-talker babble was presented at 62 dB SPL.

The Build-a-Sentence test (BAS) (Tye-Murray et al., 2008) was used to assess if the benefits of training would transfer to material and activities and a talker that were not included in the training. All items were spoken by an unfamiliar female talker. The BAS is a closed set sentence-level word test. The BAS test is comprised of 36 nouns that are selected randomly without replacement and placed in one of four possible sentence contexts (e.g., *The wife and the bear watched the mouse.; The team and the mouse watched the girls and the dog.; The boys watched the whale and the snail.; The geese watched the saint.*). Participants were asked to repeat the sentence aloud after each presentation. A list of the 36 possible words was displayed orthographically on the monitor in the test booth. Scoring was based on the number of target words identified. Six sentences (18 words) were presented in six-talker babble at six different SBRs ranging from -10 to 15. Scores at the extremes were at or near floor and ceiling, therefore, only data for the middle SBRs (-5, 0 and 5) are discussed.

RESULTS

Figure 2 depicts results for each of the four TAP tests. Table 1 presents difference scores for each test, computed between the Pre, Post, and the 3 Mos test sessions. As Figure 2 indicates, auditory training led to improved performance on all of the TAP tasks: word identification, four-choice discrimination, fill-in-the-blank, and the sentence context tasks. Most gains appeared to have been maintained at 3 months following the end of training. Table 1 indicates that the difference scores between the Pre and the Post test sessions were similar for the two training groups, as were the difference scores between the Pre and the 3 Mos test sessions. A two-way mixed design repeated measures ANOVA was performed for each of the TAP tests shown in Figure 2, with test time (Pre, Post, 3 Mos) entered as the within-subjects variables and training group entered as a between-subjects variable. Each of the four TAP assessments, word identification, four-choice discrimination, fill-in-the-blank, and the sentence context task indicated a significant main effect for test time (Fs (2, 90) = 58.69, 30.89, 11.13, and 8.89, all *p*s $\leq$ .001, PEtSqs = .566, .407, .198, and .165, respectively) and all effects for training group were non significant (Fs (1, 45) = 1.42, 2.80, 1.78, and 1.33, *p*s = .24, .10, .19, .26, PEtSqs = .031, .059, .038, .029, respectively). There were no interaction effects for any test (Fs (2, 90) = 0.68, 0.67, 0.26, and 0.14 respectively, *p*s = .76, .87, .50, and .51, respectively). These findings indicate that training led to significant improvement on all tests and that its effectiveness was comparable, regardless of whether training was provided in a massed or spaced format. Hence, there was no evidence of a distributed practice effect either in terms of overall benefit received from auditory training or in terms of retention of benefits over time following cessation of training.

---Insert Figure 2 Here---

---Insert Table 1 Here---

The difference scores between the Pre and Post test sessions and between the Pre test and the 3 Mos test sessions appear in Table 1. To further examine whether retention of benefits was evident from the Post training test session to the 3 Mos test session, a series of planned comparisons were conducted between the amount of gain at the Post assessment and the amount of retention at the 3 Mos session for each training group. The Massed training group, but not the Spaced training group, maintained a significant retention of benefits in two of the TAP-style assessments: the Fill-in-the-Blank and the Sentence Context assessment. All other indices of gain and retention for the TAP tests indicated a significant difference (Massed, all *t*s (23) > 2.2, *p*s < .05; Spaced, all *t*s (23) > 2.6, *p*s < .05; 95% CI of the Differences are provided in Table 1).

Results from the BAS testing were analyzed to determine whether the practice schedule of massed versus spaced auditory training affected the generalizability of training benefits. The results from the Pre and Post test sessions for the BAS scores appear in Figure 3. As shown in

Table 1, both training groups showed an average gain of 5.1 percentage points after training. A two-way mixed design repeated measures ANOVA revealed a significant effect of test time ($F_{(1, 45)} = 15.5$, $p < .001$, PEtSq = .257), confirming that auditory training led to significant improvement in performance on the BAS. There was no interaction effect ($F_{(1, 45)} = .001$, $p = .994$), indicating that there was not a distributed practice effect.

---Insert Figure 3 Here---

DISCUSSION

In this investigation, the major issue addressed was whether auditory training is subject to a distributed practice effect. Participants received 16 hours of auditory training. Those assigned to the Massed training group trained for five days a week whereas those in the Spaced training group trained twice a week. Our results indicate that there was no difference in overall benefit afforded by spaced versus mass training and both types of training provided benefit. Moreover, our results indicate that spaced training did not lead to better retention of benefits or to better generalization of benefits. In sum, when the amount and quality of training is equal, it appears not to matter whether patients receive auditory training five days a week or twice a week.

Of course, it is possible that training schedules positioned at the extremes of a spaced-massed training continuum might lead to different conclusions. For example, if a learner receives auditory training once a month for 16 months in a spaced practice schedule or for 8 hours per day for two consecutive days for a massed practice schedule, training effectiveness might diminish. A patient who trains for 8 hours may experience fatigue and boredom and may mentally disengage from the training activity and not accrue training benefits. The so-called *inattention theory* that has been proffered as one account for distributed practice effects (Hintzman, 1974) suggests that in massed training, an individual may pay less attention to stimuli because the training items and tasks have become increasingly familiar. The effectiveness (and ineffectiveness) of other training schedules besides those considered in this investigation might be considered in future studies.

From a larger theoretical perspective, the lack of positive effect of spaced training observed in this study suggests that the spacing effect, which has been demonstrated previously in so many other areas of human learning, does not appear to generalize to the area of speech perception and, quite possibly, other areas of perceptual learning.

Although it remains unclear as to why no differences were observed between spaced and massed training, one possibility is that participants in each of the two groups benefitted from the massed and spaced components of auditory training for different reasons, leading to an overall equivalence in benefit. For the spaced training group, the benefits are consistent with a long line of research in human learning and memory demonstrating benefits of distributed practice, as reviewed in the introduction (e.g., Green & Bavelier, 2008). Regarding the massed training group, on the other hand, one distinction between the current study and previous literature on massed versus spaced learning is the degree of novelty of the task given that speech perception is a highly practiced task. It may be that with highly practiced tasks, massed training may be more effective due to the need for uninterrupted exposure and feedback. For example, consider a case in which one might be learning to navigate using eyeglasses that inverted the visual field (see Slater, 1998, for an overview of the task). It is not difficult to imagine why in such a situation that the individual would benefit more from massed practice than from spaced practice, the latter of which may not provide the type of uninterrupted continuous exposure needed to adapt to the distorted input. One interesting line for future research would be to compare massed versus spaced training on other highly familiar tasks, such as lipreading and speechreading.

We note that another possibility is that the current study was underpowered to detect differences between the massed and spaced training regimens. Ideally, one would conduct a power analysis a priori to determine appropriate sample size. However, it is unclear what effect size should be used in conducting such an analysis for the current study because there are no published data that directly compare two methods of auditory training using the same stimuli. Insufficient statistical power is an unlikely explanation for the absence of differences between the two groups for several reasons. First, the total sample size used in the current study is larger than most studies of computerized auditory training. For example, the current sample size is larger than all but one of the 13 studies of computerized auditory training evaluated by Henshaw and Ferguson (2013). Second, although cross-domain comparisons need to made cautiously, significant differences between massed and spaced training have been observed with sample sizes less than in the present study. For example, McDaniel, Fadler, and Pasher (2013) reported significant differences between massed and spaced training on function learning with a sample size of 20 participants (compared with 24 and 23 participants for massed and spaced, respectively in the current study). Therefore, although insufficient statistical power is always a concern with findings of no differences between manipulations, in comparison with other studies, the current investigation had a relatively large sample size.

Another possibility is that the lack of benefit of spacing could be tied to the participants in this study being older, as opposed to younger adults. Given previous research demonstrating benefits for spacing over massed training in both younger and older adult populations (see, e.g., Kornell et al., 2010, for an example and a review of previous studies), from our perspective, it is unlikely that the null effect observed in this study is due to age. Nevertheless, this possible explanation cannot be completely ruled out, especially if one considers that the magnitude of the benefits of spacing can diminish among older adults relative to benefits of spacing among older adults (Simone, Bell & Cepeda, 2012), pointing to the need for (and value of) one or more future studies on spacing versus massed auditory training among younger adults or among both younger and older adults.

Overall, auditory training led to significant improvement in speech recognition, underscoring the efficacy of meaning-oriented auditory training for adults who have hearing loss. Based on their recent meta-analysis, Henshaw and Ferguson (2013) suggested that the efficacy of computerized auditory training for adults is “not robust” and that there was a paucity of evidence-based studies to assess efficacy (see also Sweetow & Palmer, 2005). The present findings add to the growing amount of evidence that training can be effective, and indeed robust (e.g., Burk, Humes, Amos, & Strauser, 2006; Sweetow & Henderson-Sabes, 2006; Barcroft et al., 2011; Tye-Murray et al., 2016; Barcroft et al., 2016).

One might argue that for at least the TAP tests in the present study participants simply learned to take a particular kind of test without improving their general abilities to recognize speech. However, in light of previous findings in this area, not only is training on the task desirable but hearing health care specialists should train patients to recognize the very speech stimuli and the very talkers that any given person with hearing loss desires. This is precisely the approach taken in several recent studies (Barcroft et al., 2011; Barcroft et al., 2016), wherein we trained persons with hearing loss to discriminate the speech of their significant others (Tye-Murray et al., 2016) with gains in a four-choice discrimination task of approximately 14% words correct. At issue here is how “generalizability” actually reflects the degree to which the component perceptual and cognitive processes engaged during training overlap with those that underlie the desired outcome when it comes to task, stimulus, and talker (see Barcroft et al., 2016).

Additionally, there are at least two more reasons to believe that generalizability was achieved in the present investigation and that speech recognition improved as a result of participation in this program of meaning-oriented auditory training. First, if improvement on the TAP tests was simply a matter of learning the test tasks, then there should have been a spacing effect since spaced training has been shown to improve performance on cognitive tasks such as working memory and negotiating a water maze better than massed training (Shebilske, Goettl, Corrington, & Day, 1999b; Sisti, Glass, & Shors, 2007). Second, auditory training also led to significant improvement on the BAS test, which, unlike for the TAP tests, was not included as a training activity. From the perspective of degree of overlap in the component processes engaged at study and at test, we propose that the combination of the processes engaged during the word-level and sentence-level activities at study overlapped to a sufficient degree with those needed for the BAS test, which is sentence-level task that is sufficiently constrained (with one overall sentence frame) so as to also reflect improvements in speech perception at the word level. This perspective on generalizability is consistent not only with the benefits of training observed on the TAP and BAS tests in the present study but also with numerous observations of lack of generalizability in numerous studies on auditory training (see Sweetow & Palmer, 2005 and Henshaw & Ferguson, 2013, for examples of lack of generalizability).

The findings have both theoretical and clinical implications. Theoretically, they suggest that, at least within the timeframes of the current study, perceptual learning might be subject to different influences than other types of learning, such as vocabulary learning. Clinically, the results suggest that clinicians can allow themselves latitude when recommending training schedules and that patients have flexibility as to whether they want to receive a prescribed number of training hours or exercises in a short amount of time or a more extended amount of time.

Acknowledgments

The first two authors are co-founders of clEAR; a Limited Liability Corporation that sells auditory training material to be provided as clinical intervention for hearing loss. We thank Elizabeth Mauzé and Shannon Sides for their contributions in data collection. This research was supported by National Institutes of Health Grant #RO1DC008964.

REFERENCES

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316-321.

Barcroft, J., Spehar, B., Tye-Murray, N., & Sommers, M. S. (2016). Task- and talker-specific gains in auditory training. *Journal of Speech, Language and Hearing Research.59,* 862-870.

Barcroft, J., Mauzé, E., Schroy, C., Tye-Murray, N., Sommers, M. & Spehar, B. (2011). Improving the quality of auditory training by making tasks meaningful. *ASHA Perspectives on Aural Rehabilitation and Its Instrumentation, 7, 1,* 15-28.

Barcroft, J., Sommers, M. S., Tye-Murray, N., Mauzé, E., Schroy, C., & Spehar, B. (2011). Tailoring auditory training to patient needs with single and multiple talkers: Transfer-appropriate gains on a four-choice discrimination test. *International Journal of Audiology*, *50*, 802-808.

Bello-Medina, P. C., Sánchez-Carrasco, L., González-Ornelas, N. R., Jeffrey, K. J., & Ramirez-Amaya, V. (2013). Differential effects of spaced vs. Massed training in long-term object-identity and object-location recognition memory. *Behavior Brain Research, 250,* 102-113.

Burk, M. H., & Humes, L. E. (2008). Effects of long-term training on aided speech-recognition performance in noise in older adults. *Journal of Speech, Language and Hearing Research*, *51*, 759-771.

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*, 263-278.

Carew, T. J., Pinsker, H. M., & Kandel, E. R. (1972). Long-term habituation of a defensive withdrawal reflex in Aplysia. *Science*, *175*, 451-454.

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

Cornoldi, C., Longoni, A. (1977). The MP-DP effect and the influence of distinct repetitions on recognition of random shapes. *Italian Journal of Psychology, 4,* 65-76.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43,* 627-634.

Ebbinghaus, H. (1885). *Über das gedächtnis: untersuchungen zur experimentellen psychologie*. Germany: Duncker & Humblot.

Folstein M.F., Folstien S.E., McHugh P.R. (1975) Minimental state: A practical method for grading the cognitive state of the patient for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Fu, Q. J., Galvin, J., Wang, X., & Nogaki, G. (2004). Effects of auditory training on adult cochlear implant patients: A preliminary report. *Cochlear Implants International*, *5*, 84-90.

Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, *23*, 692-701.

Henshaw, H., & Ferguson, M. A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence. *PLOS ONE* 8(5): e62836.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium* (pp. 77-97). Potomac, MD: Erlbaum.

Humes, L. E., Burk, M. H., Strauser, L. E., & Kinney, D. L. (2009). Development and efficacy of a frequent-word auditory training protocol for older adults with impaired hearing. Ear & Hearing, 30(5):613-27. doi: 10.1097/AUD.0b013e3181b00d90.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and indication in younger and older adults. *Psychology and Aging, 25, 2,* 498-503.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, *59*, 277-287.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 33,467-476.

Melton, A. W. (1967). Repetition and retrieval from memory. *Science*, *158*, 532-532.

Molloy, K., Moore, D. R., Sohoglu, E., & Amitay, S. (2012). Less is more: Latent learning is maximized by shorter training sessions in auditory perceptual learning. *PloS One*, *7*, e36929.

Moss, V. D. (1996). *The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student* (Doctoral dissertation, Utah State University, 1995). *Dissertation Abstracts International,* 56, 5204.

Oba, S. I., Fu, Q. J., & Galvin III, J. J. (2011). Digit training in noise can improve cochlear implant users' speech understanding in noise. *Ear and Hearing*, *32,* 573-581.

Russo, R., Parkin, A., Taylor, S., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 819-829.

Shebilske, W. L., Goettl, B. P., Corrington, K., & Day, E. A. (1999). Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology: Applied*, *5*, 413-437.

Simone, P. M., Bell, M. C., & Cepeda, N. J. (2012). Diminished but not forgotten: Effects of aging on magnitude of spacing effect benefits. *Journal of Gerontology Series B: Psychological and Social Sciences*, doi:10.1093/geron/gbs096

Sisti, H. M., Glass, A. L., & Shors, T. J. (2007). Neurogenesis and the spacing effect: Learning over time enhances memory and the survival of new neurons. *Learning and Memory*, *14*, 368-375.

Slater, A. (1998). *Perceptual Development: Visual, Auditory, and Speech Perception in Infancy.* East Sussex, Eng: Psychology Press.

Stacey, P. C., Raine, C. H., O'Donoghue, G. M., Tapper, L., Twomey, T., & Summerfield, A. Q. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology*, *49*, 347-356.

Stecker, G. C., Bowman, G. A., Yund, E. W., Herron, T. J., Roup, C. M., & Woods, D. L. (2006). Perceptual training improves syllable identification in new and experienced hearing aid users. *Journal of Rehabilitation Research and Development*, *43,* 537-552.

Stickgold, R., James, L., & Hobson, A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience, 3,* 1237-1238.

Sweetow, R. W., & Henderson-Sabes, J. (2006). The need for and development of an adaptive listening and communication enhancement (LACE™) program. *Journal of the American Academy of Audiology, 17*, 538-558.

Sweetow, R., & Palmer, C. V. (2005). Efficacy of individual auditory training in adults: A systematic review of the evidence. *Journal of the American Academy of Audiology*, *16,* 494-504.

Tye-Murray, N., Sommers, M., & Barcroft, J. (2011). I Hear What You Mean: The state of the science in auditory training. *ENT and Audiology News, 20, 4,* 84-86.

Tye-Murray, N., Sommers, M., Mauzé, E., Schroy, C., Barcroft, J., Spehar, B. (2012). Using patient perceptions of relative benefit and enjoyment to assess auditory training. *Journal of the American Academy of Audiology, 23, 8,* 623-624.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., & Rose, N. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology, 47 (S2),* S31-S37

Tye-Murray, N., Spehar, B., Sommers, M. S., & Barcroft, J. (2016). Auditory training with frequent communication partners. *Journal of Speech, Language and Hearing Research. 59,* 871-875.

Tye-Murray, N. (2015). *Foundations of Aural Rehabilitation: Children, Adults, and Their Family Members*. Stamford, CT: Cengage Learning.

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: allowing children time to forget promotes their ability to learn. *Child Development Perspectives, 8, 3,* 163–168.

Wang, Z., Zhou, R., & Shah, P. (2014). Spaced cognitive training promotes training transfer. *Frontiers in Human Neuroscience, 8, 217.*

Zhang, T., Dorman, M. F., Fu, Q. J., & Spahr, A. J. (2012). Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation. *Ear and Hearing, 33,* e70-e79.

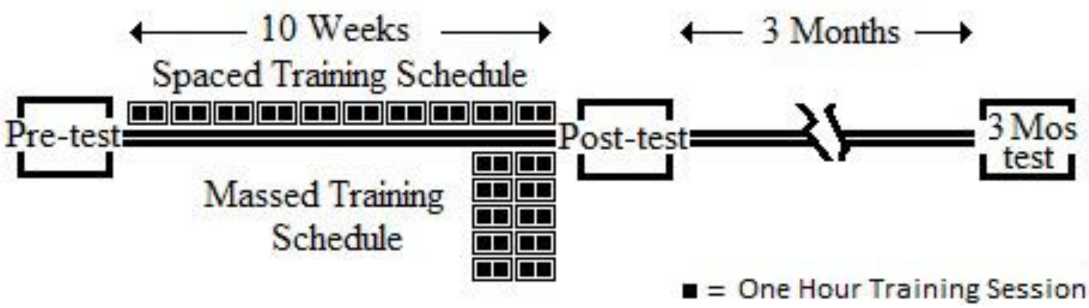
Table 1. Average (and Standard Deviation) for Pre to Post training gain scores and for Pre to 3 Mos Retention scores along with the confidence interval for each of the outcome measures split by experimental group. Pre to Post Gain is the difference between the pre training and post training outcome measures. Pre to 3 Mos Retention is the difference between the pre training and three-months-post training outcome measures.

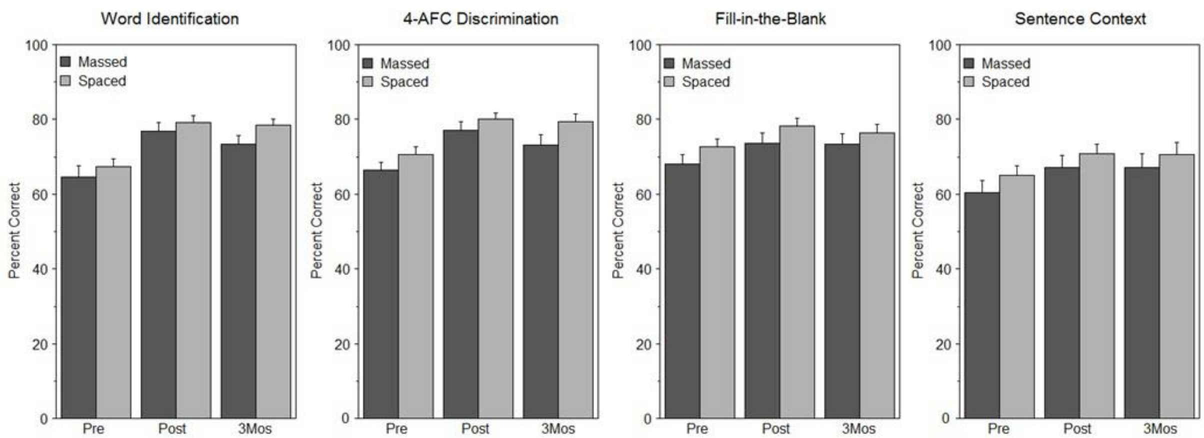
Figure 1. Schematic of training schedule for the Massed and Spaced training groups.

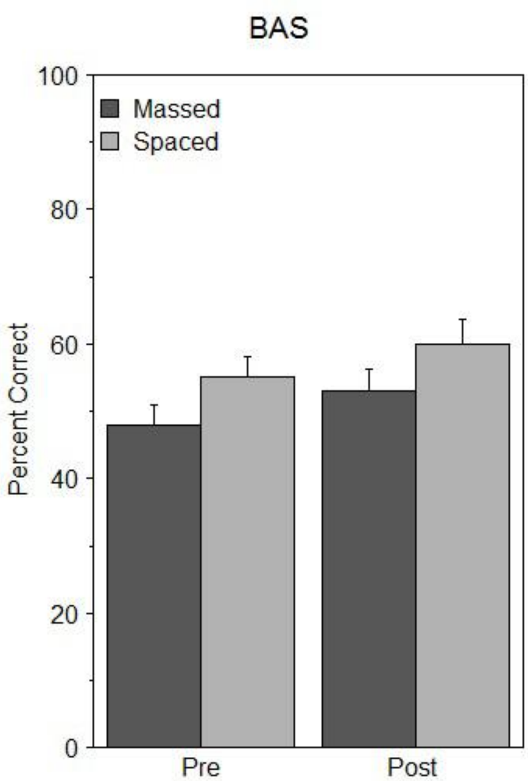
Figure 2. Test scores for the four TAP tests for the three test sessions collected from participants in the massed and the spaced training groups. Error bars indicate standard error.

Figure 3. Test scores for the BAS test for the three test sessions, collected from participants in the massed and the spaced training groups. Error bars indicate standard error.

	Group	Pre to Post Gain	95% CI of the Difference Gain	Pre to 3 Mos Retention	95% CI of the Difference Retention
Word Identification	Massed	12.2 (10)	7.9 – 16.5	8.9 (8)	5.4 – 12.3
	Spaced	11.6 (8.3)	8.1 – 15.1	10.9 (7.9)	7.6 – 14.3
4 AFC Discrimination	Massed	10.5 (9.2)	6.6 - 14.5	6.7 (10.3)	2.3 - 11.2
	Spaced	9.6 (9.1)	5.7 – 13.4	8.9 (9.8)	4.7 – 13.0
Fill in the Blank	Massed	5.5 (6.9)	2.6 – 8.5	5.3 (7.1)	2.3 – 8.4
	Spaced	5.7 (10.4)	1.4 – 10.1	3.9 (11.5)	-.99 – 8.7
Sentence Context	Massed	6.6 (11.8)	1.5 – 11.7	6.6 (13.9)	.62 – 12.6
	Spaced	5.3 (9.8)	1.2 – 9.5	5 (12.2)	-.13 – 10.1
BAS	Massed	5.1 (8.6)	1.4 – 8.8		
	Spaced	5.1 (9.1)	1.2 – 8.9		


10 Weeks
3 Months
Spaced Training Schedule
Pre-test
Post-test
3 Mos test
Massed Training Schedule
■ = One Hour Training Session


Word Identification
4-AFC Discrimination
Fill-in-the-Blank
Sentence Context
Massed
Spaced
Percent Correct
0
20
40
60
80
100
Pre
Post
3Mos


BAS
Massed
Spaced
Percent Correct
0
20
40
60
80
100
Pre
Post