Research Forum

Task- and Talker-Specific Gains in Auditory Training

Joe Barcroft,[a] Brent Spehar,[b] Nancy Tye-Murray,[b] and Mitchell Sommers[a]

Purpose: This investigation focused on generalization of outcomes for auditory training by examining the effects of task and/or talker overlap between training and at test.
Method: Adults with hearing loss completed 12 hr of meaning-oriented auditory training and were placed in a group that trained on either multiple talkers or a single talker. A control group also completed 12 hr of training in American Sign Language. The experimental group's training included a 4-choice discrimination task but not an open-set sentence test. The assessment phase included the same 4-choice discrimination task and an open-set sentence test, the Iowa Sentences Test (Tyler, Preece, & Tye-Murray, 1986).

Results: Improvement on 4-choice discrimination was observed in the experimental group as compared with the control group. Gains were (a) highest when the task and talker were the same between training and assessment; (b) second highest when the task was the same but the talker only partially so; and (c) third highest when task and talker were different.
Conclusions: The findings support applications of transfer-appropriate processing to auditory training and favor tailoring programs toward the specific needs of the individuals being trained for tasks, talkers, and perhaps, for stimuli, in addition to other factors.

*A*uditory training refers to training that is directed toward improving the ability to perform tasks related to audition, such as when individuals with hearing loss train to improve their ability to recognize and comprehend speech using their residual hearing. Methods of auditory training for individuals with hearing loss (hereafter referred to as *auditory training*) can vary on multiple fronts, such as their degree of difficulty; the frequency of the training; the overall time spent training; the extent to which they are oriented toward meaning versus form (the latter involving little or no focus on meaning); the tasks they use during training sessions, including the degree of variability in tasks; and the talkers selected to produce the spoken stimuli during training, including the degree of talker variability (see Tye-Murray, 2015, for an overview of auditory training methods).

Traditional auditory training typically involves activities that are both *analytic* (listeners focus their attention on fine-grained, acoustic differences between the elements composing spoken communication) and *synthetic* (listeners

learn to understand the meaning of an utterance, even if they do not understand every word), using live voice in a clinical setting with clear speech and good diction that are not representative of the real world. This type of auditory training can also be very drill-based, repetitive, boring, and not relevant to real-world situations. For example, a *discrimination* auditory training exercise typically presents two syllables, such as *pop* and *bob,* and the student's task is to determine whether the two syllables are the *same* or *different* (e.g., Stout & Windle, 1992). Not only does this task fail to engage most students, but it lacks ecological validity because an individual rarely, if ever, has to decide whether two consecutive words differ. In contrast, the more recent, computer-based auditory training program *Customized Learning: Exercises for Aural Rehabilitation* (*cl*EAR; previously named *I Hear What You Mean*; see, e.g., Tye-Murray, Sommers & Barcroft, 2011) now includes (a) a variety of meaning-oriented activities in which clients map word-, sentence-, and discourse-level phonological forms to the meaning or information they convey; (b) variations in talkers and the number of talkers who produce stimuli to which clients are exposed, including the possibility of using stimuli produced by the frequent communication partners of clients (see Tye-Murray, Spehar, Sommers, & Barcroft, 2016); (c) variations in signal-to-noise ratios (SNRs) that target specific performance levels (e.g., 80% correct), which can make training more focused and challenging and more

[a]Washington University, St. Louis
[b]Washington University, St. Louis School of Medicine
Correspondence to Joe Barcroft: barcroft@wustl.edu

representative of the real world; and (d) the use of games and gamelike activities designed to engage clients to a greater extent during training.

How Effective Is Auditory Training?

Systematic reviews of research on auditory training (Henshaw & Ferguson, 2013; Sweetow & Palmer, 2005) have led to disappointing conclusions when it comes to (a) the overall quality of research that has been conducted to date and (b) the overall available evidence in favor of the efficacy of auditory training. Sweetow and Palmer (2005) found, for example, that only 6 of 213 articles provided sufficient evidence-based assessments of the benefits of auditory training, and of the six articles that did provide that information, only four suggested it was beneficial, one did not, and one presented mixed results. In a more recent meta-analysis of research on computer-based auditory training, Henshaw and Ferguson (2013) found that only 13 of 229 articles published since 1995 met their methodological criteria for inclusion. It is somewhat encouraging that 9 of the 10 articles assessing on-task learning ("any improvement in performance on a task or stimulus that had been directly trained," p. 5) did find improved on-task learning performance. In other words, despite the discouraging nature of the overall picture presented by the most recent research reviews, demonstrations of improvement in on-task learning do appear to be the norm in most auditory training studies deemed to be of sufficient methodological rigor. Improvement of this nature satisfies what Dubno (2013, p. 2) deemed a minimum expectation for auditory training: "At a minimum, it is expected that significant improvements will be observed [in auditory training] for tasks (e.g., closed-set recognition) and stimuli (e.g., monosyllabic words in isolation spoken by a single talker) that are used during training."

Whereas on-task learning appears to be obtainable with a substantial degree of reliability, unfortunately, as the Henshaw and Ferguson (2013) descriptive summary (in their table 3) indicates, there is little evidence for generalization of the benefits of auditory training beyond on-task training and assessment. When performance is assessed using one or more tasks other than those performed during training, little or no benefit is observed. This overall pattern of findings suggests that the efficacy of auditory training is largely task specific, that is, tied to, or dependent upon, the degree to which the task performed during training and at testing are similar or the same. Other evidence suggests that auditory training may also be largely talker specific, such as when single-talker–trained participants perform particularly well when tested with stimuli produced by the same talker (see Barcroft et al., 2011). Finally, the effects of training can also be stimulus specific or tied to the specific stimuli with which one is trained. Burk and Humes (2008), for example, found that auditory training led to improved word recognition for trained words but no improvement for recognition of untrained words. Of course, in addition to each of these types of specificity (task, talker, and stimulus), benefits also may be affected by other factors, such as the nature of the training regimen, including "dosage" (see Humes, Kinney, Brown, Kiener, & Quigley, 2014), which is another variable to consider when predicting the probable learning outcomes of auditory training programs.

Examples of Task-, Talker-, and Stimulus-Specific Benefits

Consistent with the roles of task-, talker-, and stimulus-specificity described above are studies that have demonstrated more pronounced auditory training benefits in transfer-appropriate contexts, that is, contexts in which the tasks and the talkers used during training and testing are the same or overlap to varying degrees. The general pattern supported is the following: the more the overlap between task, talker, and stimulus at study and at test, the more likely it is for auditory training benefits to be observed. With regard to task-specific effects, for example, Burk and Humes (2008) demonstrated benefits on word recognition when training and testing involved individual words but not when training involved individual words and testing involved sentences. Regarding talker-specific effects, Barcroft et al. (2011) found that participants trained in a single-talker condition improved significantly more on a four-choice discrimination task (4AFC) when their posttraining test was produced by the same talker with whom they had trained. Lastly, for stimulus-specific effects, which concern not only the actual stimuli used in training but also the task in which they appear, Burk, Humes, Amos, and Strauser (2006) found that word-recognition improvements went from 45.3% for trained stimuli to 6.9% for untrained stimuli for an open-set context, and from 11.0% for trained stimuli to no improvement for untrained stimuli in a closed-set context. These are just a small sampling of findings that point to the pivotal role of task-, talker-, and stimuli-specific benefits in auditory training.

Degree of Overlap and Task-, Talker-, and Stimulus-Specific Effects

Findings such as these illustrate the need for an overall transfer-appropriate view of auditory training, one in which increased overlap in task, talker, and stimulus from training to assessment is expected to produce increased gains. Figure 1 visually depicts this perspective in a Venn diagram. The more two or all three of the circles in the diagram overlap, the greater the gains that one should expect. Clearly, additional factors, such as type of background noise, can be added to the schematic, but for present purposes, Figure 1 depicts our general view on how increasing the degree of overlap in task, talker, and stimulus should increase gains observed from auditory training. As we increase the extent to which tasks, talkers, and stimuli used at training and test are the same, we expect the observable benefits of training to increase as well. Although this perspective does not deny generalization as an appropriate goal of

Figure 1. Schematic of the degree of overlap (in task, talker, and stimulus) in auditory training.
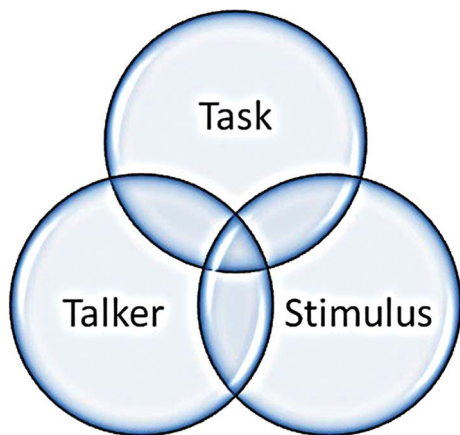
training, it offers a means of addressing the issue of generalization more directly and teasing apart specific aspects of generalizability (and lack thereof) in this area.

The Present Study

The data reported in this research forum article provide a unique opportunity to assess the relationship between degree of overlap in task and talker during training and at testing. The participants in the study were trained for 12 hr for a number of weeks using the *cl*EAR program. Participants were either trained using one talker (Single-Talker training) or, using the same training battery, with six talkers (Multi-Talker training). One of the activities they completed during each training session was a four-choice discrimination task (described below) with either a single talker or multiple talkers producing the stimuli. At test, they were asked to perform the same task using stimuli produced by the assigned single talker, the same set of multiple talkers, and by a novel talker. This study–test format provided an opportunity to assess learning performance when the degree of overlap (at study and at test) in the task was 100%, whereas the degree of overlap in the talker was either 100% or less than 100%, such as when a participant in the Multi-Talker group was tested in the Single-Talker format, which included only one of the six talkers that the participant had heard during training and when both groups were tested using a novel talker. The same participants also completed sentence-level training activities as part of the *cl*EAR training program in either Multiple-Talker or Single-Talker formats, but they were not tested using the same sentence-level training activities. Instead, they completed the Iowa Sentences Test (Tyler, Preece, & Tye-Murray, 1986; described below) as an assessment measure, providing an opportunity to assess the effects of a degree of overlap at substantially less than 100% overlap (between study and at test) and well below that of the study–test overlap in the case of the four-choice discrimination training and testing.

The Four-Choice Discrimination Task and the Iowa Sentences Test

The four-choice discrimination task is an analytic task that required participants to choose among four pairs of pictures on a screen that best matched the word pair that was presented auditorily. For example, the pair *shower–flower* could be presented auditorily, and participants viewed pictures that corresponded to the four options *flower–shower, flower–flower, shower–flower,* or *shower–shower*. Note that a discrimination task of this nature, in contrast to a same–different, syllable-oriented task, requires attention to meaning (the meaning or referents of targets words must be accessed to arrive at correct responses), making this task more ecologically valid than a synonym–discrimination task in the real world. If they selected, *shower–flower,* their response was correct. The Iowa Sentences Test, on the other hand, was a task that included 20 items (selected from the full version of the Iowa Sentences Test). Participants were not trained on this task during the training phase, and all of the items on the test were produced by different talkers and by novel talkers.

Assessing Different Degrees of Overlap

For all participants, we first report on the pretraining test (Pre-test) to posttraining test (Posttest) results for the Iowa Sentences Test. Because degree of overlap between the tasks performed during training and the Iowa Sentences Test was much less than 100%, little or no gains would be expected when participants were tested using this task. The Iowa Sentences Test did not include any of the same sentences on which the Multi-Talker and Single-Talker participants were trained. It also did not include any of the same speakers with whom participants trained. Therefore, the degree of overlap in task between training and testing for both groups was either minimal or none. If one considers the sentence-level training that participants received (see Activities 3 and 4 below) to be at least slightly similar to the Iowa Sentences Test, then the overlap was minimal but not zero. If one considers the sentence-level training that participants received to have overlapped slightly with the Iowa Sentences Test, then the overlap was not zero but was still minimal when it comes to task and talker.

Finally, after reporting the Iowa Sentences Test data, we report Pretest to Posttest data on the four-choice discrimination task for which talker at study and at test overlapped 100% in the Single-Talker group but only 16.7% (1 out of 6 talkers) in the Multi-Talker group, and even less so for both groups when testing was done with a novel talker. Therefore, (a) the third highest degree of overlap corresponded to the four-choice discrimination task and both the Single-Talker and Multiple-Talker training, (b) the second highest degree of overlap corresponded to four-choice discrimination and Multiple-Talker training, and (c) the highest degree of overlap corresponded to the four-choice discrimination task and Single-Talker training. Given these conditions related to degree of overlap, the expected

observable gains move from lowest to highest in each of the following four conditions: (a) the Control participants who received no auditory training, (b) the trained participants in both Single-Talker and Multi-Talker groups when tested using the Iowa Sentences task, (c) both trained groups on the Novel-Talker four-choice Posttest, (d) the Multi-Talker participants on the Single-Talker test, and (e) the Single-Talker–trained participants on the Single-Talker test.

Research Questions

The present analyses were guided by the following research questions:

1. Does meaning-oriented auditory training with the *cl*EAR program lead to gains in performance on a four-choice discrimination task and/or the Iowa Sentences Test?

2. To what extent are gains dependent on the degree of overlap in the task performed during training and at testing?

3. To what extent are gains dependent on the degree of overlap in the talker(s) who produce stimuli during training and at testing?

4. Are gains the highest when both task and talker(s) overlap to the highest degree?

Method

Participants

One-hundred and seven adults with hearing impairment served as participants. Forty-one participants were assigned to the Multi-Talker–trained group (21 women; 20 men) (age, $M = 66.3$, $SD = 18.2$); 42 participants (17 women; 25 men) (age, $M = 68.5$, $SD = 13.8$) were assigned to the Single-Talker–trained group, and the Control group had 24 participants (12 women; 12 men) (age, $M = 66.0$, $SD = 11.1$). Participants were assigned to the Multi-Talker–trained and Single-Talker–trained groups in an alternating manner (i.e., every other recruit was assigned to the Multi-Talker–trained group). Those in the Control group were recruited and assigned as a single block. Results from 69 of the participants who were assigned to a training group were reported previously in Barcroft et al. (2011). There were no significant differences among the mean ages of the three groups, $F(2, 104) < 1$, $p > .05$. Pure-tone averages in dB HL (PTA dB HL; average unaided threshold for 0.5, 1, and 2 kHz pure tones) of the better ear for the Single-Talker–trained, Multi-Talker–trained, and the Control groups were 48.8 ($SD = 16.9$), 51.0 ($SD = 14.1$), and 41.7 ($SD = 14.5$) dB HL, respectively. There were no significant differences among the mean PTAs of the better ear for the groups, $F(2, 104) = 2.87$, $p > .05$. Aided speech perception for single words in a carrier phrase in quiet was measured using the NU-6 word test (Tillman & Carhart, 1966). Scores for the Single-Talker–trained, Multi-Talker–trained, and the Control groups were 74.2% ($SD = 22.1$), 66.8%

($SD = 23.4$), and 84.1% ($SD = 11.9$), respectively. Despite the overall lack of differences in PTA among the groups, a difference in aided speech perception in quiet for the NU-6 word test was noted, $F(2, 104) = 5.2$, $p < .01$, $\eta_p^2 = .09$). Further analyses indicated that this difference was due to the Control group having significantly higher speech-perception scores than the Multi-Talker–trained group ($p < .01$, 95% confidence intervals [CIs] of the difference [4.2, 30.3]) whereas all other comparisons were not significantly different (all $p > .05$). It should be noted that all comparisons in the current study were based on within-participant changes from before to after training and, as long as floor and ceiling conditions do not exist, these comparisons are still appropriate despite the differences in starting points for the NU-6 in quiet among the groups.

All participants were native English-speaking, community-dwelling residents recruited through the Volunteers for Health at Washington University School of Medicine and received $10/hr for their participation. None reported having ever participated in lipreading or speech-reading training. Participants were screened to exclude those who had had neurologic events, such as stroke or open or closed head injuries. To screen for dementia, participants completed the Mini Mental Status Exam (Folstein, Folstein, & McHugh, 1975). Individuals who scored below 24 (of a possible 30) were excluded from the study.

Training

Participants who received auditory training were required to return to the speech and hearing laboratories at Washington University in St Louis twice per week for 6 weeks. Each session lasted approximately 1 hr. During the training, participants were seated approximately 0.5 m from a touch screen in a sound-treated room. Audio presentations were provided through two loudspeakers positioned at ±45° to the participant's forward position. All activities (except Activity 5; see below) were conducted in six-talker babble, presented at approximately 62 dB SPL. The level of the target speech was adapted to the listener's responses using a two-down, one-up procedure to keep performance at approximately 71% (Levitt, 1970). Responses were made via the touch screen, and the next presentation did not occur until a response was made. When participants were incorrect, the wrong answer would disappear from the screen and the trial would be repeated at an SNR that was +2 dB easier. The next item was not presented until a correct response to the current trial was made. The decision to make the next trial harder or easier was based only on the first response to the previous trial.

Each visit consisted of training using all five activities. No single trial or set of stimuli was repeated throughout the twelve 1-hr sessions. A detailed description of each activity is provided in Tye-Murray et al. (2012), and a brief description of each activity is presented here. Each activity took approximately 10–12 min to complete. Activity 1 was an analytic task that required participants to determine whether a target sound (i.e., /t/) was in an initial, medial,

or final position of a presented word (e.g., *team, center,* or *fight,* respectively). Activity 2, also analytic, required participants to choose among four pictures on the screen that best matched the word pair that was presented auditorily (e.g., *flower–shower, flower–flower, shower–flower,* or *shower–shower*). Activity 3 was an analytic/synthetic, fill-in-the-blank task in which participants chose, among four alternatives presented auditorily, the best final word of a sentence that was also provided auditorily (e.g., Twenty percent is a generous (*nip, tip, dip, kip*). Activity 4, also an analytic/synthetic task, required participants to choose the best of three sentences that appeared on the screen to go with the sentence that was presented auditorily. For example, the sentence *The student had a broken leg* was presented auditorily, and the participant chose among options on the screen, including *He had to sharpen his pencil during the test, He had to sit with his foot elevated,* and *His glasses needed to be repaired.* The options appeared as text, and the participant selected a response by pressing a virtual button on a touch screen. Finally, Activity 5 was a discourse-comprehension task in which participants listened to a short, 2–3 min paragraph, in a constant SNR of +2 dB and answered questions based on the content of the passage. All participants received the same training items in the same order. The Single-Talker–trained group heard items from one of the six recorded talkers. Talker assignment in the Single-Talker–trained group was counterbalanced. The Multi-Talker–trained group received items from all talkers in a counterbalanced order.

Participants assigned to the Control group were also asked to come to the labs twice a week for 12 weeks. However, instead of auditory training, they received sign language instruction for one hour each session. During their visits to the clinic, they progressed at their own pace through the American Sign Language (ASL) lessons at www.lifeprint.com (Vicars, n.d.). These ASL lessons were selected for the Control group in consideration of how they involve linguistic content without audition. As such, the ASL lessons could be engaging but should not improve the participants' ability to make use of their residual hearing. Note that participants who received auditory training may have had greater confidence than the ASL (Control) participants on at least some of the posttests given that activities in which they engaged during training (e.g., the 4-choice discrimination task; see description below) were more like the activities in which they engaged on the posttests relative to what was the case with the Control (ASL) participants.

Outcome Measures

In the current investigation, we report findings for two outcome measures selected from a larger assessment battery. Assessments from the larger test battery also included tests of discourse comprehension, consonant-level perception, and unaided speech discrimination. The assessments reported here were specifically chosen to contrast items using the transfer-appropriate processing framework (Morris, Bransford & Franks, 1977). According to this framework, the memorial effect of any type of task depends on the nature of the task at study and at test, such that, for example, a semantically oriented task at study should improve memory for a semantically oriented task at testing, and the same would hold true for structural tasks and other types of tasks. In this sense, learning can be posited to be transfer-appropriate. Both outcome measures were presented in the same noise used during training at an SNR of +2dB.

Iowa Sentences Test

The first assessment includes 20 items selected from the Iowa Sentences Test (Tyler et al., 1986) and represents the assessment items that were least transfer appropriate (different in nature at testing as compared with their auditory training battery) because none of the speakers or stimuli were presented during training. Twenty sentences, among the 100 available sentences in the Iowa Sentences Test, were selected to create each list. Each sentence within a list was spoken by a different speaker (10 men). The same list was given to all participants within each time interval (before and after), but different lists were presented at the two time intervals. Items were scored by keyword correct (i.e., articles and other function words were not considered in the scoring).

Four-Choice Discrimination

The test assessment task, which is the same as Activity 2, provided a higher degree of overlap in task from training to assessment. This task, which is termed the four-choice discrimination test, was presented in an SNR of +2 dB. Three conditions were tested within this task (Single-Talker test, Multi-Talker test, and Novel-Talker test). The conditions were presented in blocks of 36 items, and the order in which conditions were tested was counterbalanced across the participant group to reduce the effects of any learning on outcomes. Also, to reduce learning effects, three practice items from each condition were presented before any of the test trials. For all three groups, the talker used during Single-Talker testing was selected from the six possible talkers. Talker selection and assignment to participants was counterbalanced and based on participation order. For the Single-Talker–trained group, the Single-Talker test was the same talker used during training. The Multi-Talker condition included an equal number of trials from all six talkers and best approximates the type of training received by the Multi-Talker–trained group. The Novel-Talker condition included a male talker that was never used in training or elsewhere in the assessment battery.

In the group with Single-Talker training and Same-Talker testing, a number of stimuli (25%) also overlapped between training and testing, lending itself to the highest degree of overlap (and potentially the greatest gains). The goal of this study was not to manipulate stimulus as a variable. We did, however, take advantage of the opportunity to analyze performance for this particular subset of stimuli (words). Because only participants within the Single-Talker group were tested with, and then trained with, a subtest of stimuli that included the same talker along with the same stimuli, the analyses were limited to this group.

Results

Iowa Sentences Test

Scores from the Iowa Sentences Test were analyzed to assess whether training improved performance on the Iowa Sentences Test, which consisted of sentences that were not included in training. Means in all three groups (Single-Talker, Multi-Talker, and Control) differed little between the Pretest and the Posttest analysis. In the Single-Talker group, means went from 39.32% ($SD = 26.76$) to 40.70% ($SD = 26.94$). In the Multi-Talker group, they went from 37.56% ($SD = 27.01$) to 36.71% ($SD = 27.01$). In the Control (ASL) group, they went from 50.96% ($SD = 23.34$) to 48.45% ($SD = 23.89$). Results of a 2 × 3 mixed-design analysis of variance (ANOVA), with test time as a repeated measures variable (Pretest and Posttest) and training group (Single-Talker, Multi-Talker, and Control) as between-participant variables revealed no significant main effects or interactions, suggesting that auditory training did not yield benefits when the task and talkers used during training did not overlap with the task and talkers used at test for either test group.

Four-Choice Discrimination Task

To assess the transfer-appropriate gain the training provided, an analysis of the outcome measure for the 4AFC testing was conducted. The improvement among the three different groups (Single-Talker, Multi-Talker, and Control) on the test scores for the three types of 4AFC scores (Same-Talker, Multi-Talker, and Novel-Talker) was compared. It showed diminishing gains as the similarity between testing material and training material become larger. Figure 2 includes the average scores for each of the groups among the three types of tests. The gain observed for each group and type of test is summarized in Table 1.

Three separate analyses, each similar to the mixed-design repeated-measures ANOVA described for the Iowa

Sentences data, were conducted for the three types of tests. The three groups were analyzed as a between-participant variable, and the Pretest and Posttest scores for each of the tests were analyzed as within-participant variables. Each of the ANOVA results is described separately.
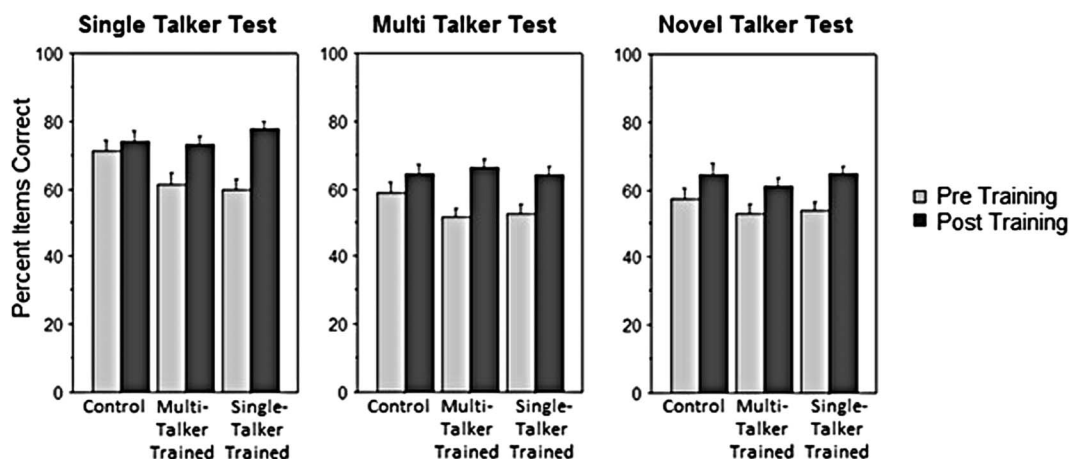
Single-Talker Test

Results of the ANOVA for the Single-Talker test indicated no significant overall main effect for Group, $F(2, 104) = 0.8$, $p > .05$; and a significant Pretest versus Posttest difference, $F(1, 104) = 117.3$, $p < .0001$, $\eta_p^2 = .53$. An interaction, however, was found between test interval and Group, $F(1.2, 104) = 17.6$, $p < .0001$, $\eta_p^2 = .25$, suggesting that the improvement between Pretest and Posttest was dependent on group. To investigate the interaction, post hoc analyses using Bonferroni-corrected t tests indicated that the Control group did not improve Pretest to Posttest on the Single-Talker 4AFC test. To further analyze the degree of overlap, a planned comparison of the gains shown by the Single-Talker versus the Multi-Talker group on the Single-Talker test was performed. The Single-Talker group improved by 6.2 percentage points more than the Multi-Talker group did. An unpaired t test indicated that the Single-Talker group improved significantly more on the Single-Talker test than did the Multi-Talker group, $t(81) = 2.67$, $p < .01$, 95% CI of the difference [1.5, 10.7]. Notably, this condition is most similar to the Single-Talker group's training with this task.

Multi-Talker Test

A similar analysis to the one conducted above was conducted for the Multi-Talker 4AFC test. Results of the ANOVA for the Multi-Talker test indicated no overall significant main effect for Group, $F(2, 104) = 0.4$, $p > .05$. A significant Pretest to Posttest difference was revealed, $F(1, 104) = 77.4$, $p < .0001$, $\eta_p^2 = .43$. A significant interaction was observed between test interval and Group

Figure 2. Pretest and Posttest training scores for each for the three groups; Control, Multi-Talker, and Single-Talker in the percentage correct on each of the three four-choice discrimination task assessments. Error bars indicate *SE*s.

Table 1. Effects of different training type by test type, presented in gain in percentage points (and *SD*).

Test	MOST OVERLAP		
	ST group	**MT group**	**Control group**
Single-Talker 4AFC test	**17.8* (10.2)**[+++]	11.6* (10.9)	2.8 (7.3)
Multi-Talker 4AFC test	**11.4* (12.5)**[++]	**14.7* (10.5)**[++]	5.7 (13.1)
Novel-Talker 4AFC test	10.7* (10.5)[+]	8.0* (12.0)[+]	7.2* (10.4)[+]
Iowa Sentences test	1.4 (10.5)	−0.9 (8.3)	−2.5 (7.5)
			LEAST OVERLAP

Note. Bolded results indicate the findings for which a significant difference between the groups could be found. Plus signs (+, ++, +++) highlight the increased degree of overlap: more plus signs reflect greater overlap. ST = Single-Talker; MT = Multi-Talker; 4AFC = four-choice discrimination task.

*Statistically significant differences.

$F(1,2,104) = 4.0$; $p < .05$, $\eta_p^2 = .07$, suggesting that the improvement Pretest to Posttest was dependent on group. To investigate the interaction, post hoc analyses using Bonferroni-corrected *t* tests indicated that the Control group did not improve Pretest to Posttest on the Multi-Talker 4AFC test.

Comparison of Single-Talker and Multi-Talker

To parallel the analysis comparing the gains for the Single-Talker and Multi-Talker groups, a planned *t* test was conducted on the 3.0 percentage points gained by the Multi-Talker group over the Single-Talker group in the Multi-Talker test. The unpaired *t* test indicated that, although the Multi-Talker group did do slightly better on the Multi-Talker test, the gains were similar for the two groups, $t(81) = 1.2$, $p > .05$. Taken with the results of the Same-Talker test, the conclusion is consistent with greater exposure to the same talker during training improving outcomes for those trained by a target talker. The two groups trained for the same amount of time. A training sample that included six different talkers with the same amount of training for the Multi-Talker group provided no improvement for any talkers as great as that found for the Single-Talker group that focused on one talker.

Novel-Talker Test

The analysis of the Novel-Talker test provided an opportunity to examine how the training might influence the ability to perceive a talker not included in the training—that is, if the task was the same as the training, but the talker was different—and whether there was improvement from Pretest to Posttest for one or more of the groups. Results of the ANOVA for the Novel-Talker test indicated no overall main effect for Group, $F(2, 104) = 0.5$, $p > .05$. A Pretest versus Posttest difference was found, $F(1, 104) = 61.2$, $p < .0001$, $\eta_p^2 = .37$, with no interaction. Consistent with results showing no interaction, the planned comparisons showed that all groups, including the untrained Control group, improved by the same amount on the Novel-Talker test. It is possible that the Control group's exposure to the task during the Pretest assessment was enough to have learned the task as well as the trained groups did. As the results pertain to the question regarding the relationship between training and testing, the finding is consistent with the

continued assertion that the more the training is different from the assessment, the worse the participants do.

Results Based on Degree of Overlap

Table 1 includes the difference in scores (and the *SD*) in percentage points for each of the groups and tests. Bolded results indicate the findings for which a significant difference between the groups could be found. Plus signs (+, ++, +++) highlight the increased degree of overlap: more plus signs reflect greater overlap. As shown, gains become greater as the number of plus signs increases, reflecting how increases in degree of overlap correspond to increases in gains.

Results Based on Talker and Stimulus Specificity

Detailed analysis of the stimulus-by-stimulus responses from the Single-Talker group's subset of responses in the 4AFC testing allowed for the comparison of performance on test items that overlapped with training (same talker and same stimuli) and those that were not included in the training (different talker and different stimuli). Average gains for test items produced by the same talker and included in the training (18.4 percentage points, $SD = 12.0$) were significantly greater than gains for test items produced by unfamiliar talkers and not included in the training (11.5 percentage points, $SD = 9.0$), $t(40) = 3.4$, $p < .01$, 95% CI of the difference [2.7, 10.9].

Discussion

The findings of this study are consistent with the general perspective (as depicted in Figure 1) that the benefits of auditory training should increase as a function of increases in the degree of overlap among task, talker, and stimulus at training and at test. From this perspective, predictions can be made about the probable amount or *level* of benefit of a given program by considering, among other factors of course, the specific tasks, talkers, and stimuli used in the training program and in the assessment battery used to test the efficacy of the training program.

Degree of Overlap and Amount of Gain

We can identify four levels of Pretest to Posttest performance in the data. First, no significant gains were observed

for any group on the Iowa Sentences Test when neither the task nor the talker used in training overlapped with task type and talker used in testing. Second, significant gains were observed for both Single-Talker and Multi-Talker groups on the four-choice discrimination task, a task that was also part of the training regime for those groups. These gains emerged regardless of the talker used when testing. In fact, even the Control group managed to achieve significant gains on one of the posttest measures (for the Novel-Talker), suggesting that even practice with the task in question on the Pretest alone was sufficient to improve scores significantly on this measure. Notably, however, the Control group did not improve Pretest to Posttest on the Iowa Sentences task whereas they did show task improvement on the 4AFC task. It is possible that this is partially due to the closed-set (4AFC) versus open-set (Iowa Sentences) nature of the tasks. The fairly similar gains in the Single-Talker– and Multi-Talker–trained groups with the Novel-Talker 4-choice testing are consistent with this level of gains as well. Third, when the task at training and at testing overlaps, and *some* of the talkers also overlap, a higher level of gain is to be expected, which is consistent with the level of gain observed for the Multi-Talker group on the four-choice task group when tested using Multiple-Talker stimuli and for the Single-Talker group when tested using Multiple-Talker stimuli. Finally, an even greater benefit is expected when training and testing involved both the same task and the same talker, which was the case for participants in the Single-Talker group on the four-choice task with the same talker. Although not systematically assessed in this study, the next-higher level in degree of overlap would be full overlap in task, talker, and stimuli from training to test, which would be expected to produce the greatest gains.

We acknowledge that comparing training effects across the Iowa Sentence and 4AFC tests can be problematic, given that performance may be tied the overall level of difficulty of each of the two tests, the amount of training and the time course of training needed to improve performance on each of the two tests (i.e., rate of learning), and other factors, such as the vocabulary used in each of two tests. Therefore, it may be, for example, that no effect of training was observed on the Iowa Sentence test because of the overall test difficulty, the open-set nature of the test, or both. Had training focused more on improving performance on the Iowa Sentence Test and not the four-choice discrimination test, more improvement might have been observed on the Iowa Sentence Test when (also) used as a posttest measure. Future studies certainly can explore this possibility and, in so doing, further refine our understanding of which specific aspects of overlap between training and testing best account for differences in gain from training.

Tailoring Auditory Training

The present findings—in combination with those of numerous previous studies showing task-, talker-, and stimulus-specific effects (e.g., Barcroft et al., 2011; Burk & Humes, 2008; Burk et al., 2006)—provide strong evidence in favor of tailoring auditory training to the specific needs of individuals with hearing loss. To the extent that we can identify specific tasks and talkers that are pertinent to the day-to-day communicative acts of an individual with hearing loss, we also can design auditory training programs that incorporate these tasks and talkers, making them more task appropriate and effective. Tye-Murray et al. (2016) have provided strong evidence of this possibility by demonstrating that adults who use hearing aids improved their perception of speech produced by their frequent communication partners. Future research can continue to expand on benefits of task-, talker-, and stimulus-specific auditory training by including other provisions, such as by focusing on task-specific challenges. If an individual with hearing loss is challenged in particular when attempting to hear directions while driving, activities can be designed to focus on hearing driving instructions (task specificity) produced by the individual's frequent communication partner when the individual is driving (talker specificity) and using lexical items that are common in this particular communicative context (stimulus specificity). In light of the task-, talker-, and stimulus-specific benefits of auditory training that have been observed to date, a training regimen of this nature would uniquely favor the day-to-day life of that individual.

Of course there are other factors, such as the nature of the background noise and access to visual information (visual cues needed for lipreading or otherwise), that can be added to the list of what affects overall processing specificity in an auditory training program. These factors also need to be considered within a larger, systematic approach that anticipates benefits across a variety of factors and possible activities that might become the focus of training. Additionally, the role of *dosage,* as mentioned previously, and other training techniques, such as perceptual fading (Jamieson & Morosan 1989), are also important considerations when designing auditory training programs that are effective and achieve the greatest generalizability possible. Nevertheless, task, talker, and stimulus are important to understanding the nature of generalizability (and lack thereof) and to customizing training based on the specific needs of clients. Greater systematic manipulation of task, talker, stimuli, and other factors in future studies can further quantify the independent and combined effects of these factors. It may be, for example, that the relative benefits from one factor (such as task) are much greater than those from another factor (such as type of background noise) or that the benefits of one factor (such as talker) do not emerge unless the benefits of another factor (such as task) have already been obtained. Future research can address these issues as well.

Conclusion

The data reviewed in the present report were uniquely suited to address the issue of generalizability (and lack thereof) when it comes to the roles of task and talker in auditory training. The findings of the study revealed a series of levels of benefits that correspond to degree of overlap in

task and talker from training to test: the greater the overlap, the larger the gains from auditory training. We provided a visual schematic (see Figure 1) that is consistent with this pattern of findings but recommend that future research continue to assess the relative effects of task-, talker-, and stimulus-specificity (among specificity related to other factors, such as type of background noise), further quantifying the relative effects of each of these in auditory training.

Acknowledgment

References

Barcroft, J., Sommers, M. S., Tye-Murray, N., Mauze, 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*(11), 802–808.

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*(3), 759–771.

Burk, M. H., Humes, L. E., Amos, N. E., & Strauser, L. E. (2006). Effect of training word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear and Hearing, 27,* 263–278.

Dubno, J. R. (2013). Benefits of auditory training for aided listening by older adults. *American Journal of Audiology, 22*(2), 335–338. doi:10.1044/1059-0889(2013/12-0080)

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

Henshaw, H., & Ferguson, M. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence. *PLoS One, 8*(5), e62836. doi:10.1371/journal.pone.0062836

Humes, L. E., Kinney, D. L., Brown, S. E., Kiener, A. L., & Quigley, T. M. (2014). The effects of dosage and duration of auditory training for older adults with hearing impairment. *The Journal of the Acoustical Society of America, 136,* 3, EL224–EL230.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques, *Canadian Journal of Psychology, 43,* 88–96.

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

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Stout, G., & Windle, J. (1992). *The developmental approach to successful listening II (DASL-II)*. Englewood, CO: Resource Point.

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*(7), 494–504.

Tillman, T. W., & Carhart, R. (1966). *An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6*. Unpublished manuscript, Northwestern University Auditory Research Laboratory, Evanston, IL.

Tye-Murray, N. (2015). *Foundations of aural rehabilitation: Children, adults, and their family members* (5th ed.). Clifton Park, NY: Delmar Cengage Learning.

Tye-Murray, N., Sommers, M. S., & 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. S., Mauze, 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), 623–634.

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

Tyler, R. S., Preece, J. P., & Tye-Murray, N. (1986). The Iowa Phoneme and Sentence Tests [laser videodisc]. Unpublished manuscript, Department of Otolaryngology—Head and Neck Surgery, University of Iowa, Iowa City, IA.

Vicars, B. (n.d.). *ASL*. Retrieved November 19, 2014, from http://www.lifeprint.com/