Article

Current Biology

Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise

Highlights

- Elderly subjects trained for 8 weeks on a computerized audiomotor interface
- Speech-in-noise intelligibility in challenging listening conditions improved by 25%
- Generalized training benefits were compared to and exceeded placebo effects
- Inhibitory control ability and game strategy predicted individual training benefits

Authors

Jonathon P. Whitton,
Kenneth E. Hancock,
Jeffrey M. Shannon, Daniel B. Polley

Correspondence

jonathon_whitton@meei.harvard.edu

In Brief

Whitton et al. put the "real-world" usefulness of computerized perceptual training to the test in a randomized, double-blind, placebo-controlled study. They report that hearing-impaired older adults can triple the speech intelligibility benefits of their hearing aids in challenging listening environments after training on a custom audiomotor game.

Whitton et al., 2017, Current Biology *27*, 1–11
November 6, 2017
https://doi.org/10.1016/j.cub.2017.09.014

CellPress

Please cite this article in press as: Whitton et al., Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise, Current Biology (2017), https://doi.org/10.1016/j.cub.2017.09.014

Current Biology
Article

Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise

Jonathon P. Whitton,[1,2,5,*] Kenneth E. Hancock,[1,3] Jeffrey M. Shannon,[4] and Daniel B. Polley[1,3]
[1]Eaton-Peabody Laboratories, Massachusetts Eye and Ear Infirmary, Boston, MA 02114, USA
[2]Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Department of Otolaryngology, Harvard Medical School, Boston, MA 02114, USA
[4]Hudson Valley Audiology Center, Pomona, NY 10970, USA
[5]Lead Contact
*Correspondence: jonathon_whitton@meei.harvard.edu
https://doi.org/10.1016/j.cub.2017.09.014

SUMMARY

Sensory and motor skills can be improved with training, but learning is often restricted to practice stimuli. As an exception, training on closed-loop (CL) sensorimotor interfaces, such as action video games and musical instruments, can impart a broad spectrum of perceptual benefits. Here we ask whether computerized CL auditory training can enhance speech understanding in levels of background noise that approximate a crowded restaurant. Elderly hearing-impaired subjects trained for 8 weeks on a CL game that, like a musical instrument, challenged them to monitor subtle deviations between predicted and actual auditory feedback as they moved their fingertip through a virtual soundscape. We performed our study as a randomized, double-blind, placebo-controlled trial by training other subjects in an auditory working-memory (WM) task. Subjects in both groups improved at their respective auditory tasks and reported comparable expectations for improved speech processing, thereby controlling for placebo effects. Whereas speech intelligibility was unchanged after WM training, subjects in the CL training group could correctly identify 25% more words in spoken sentences or digit sequences presented in high levels of background noise. Numerically, CL audiomotor training provided more than three times the benefit of our subjects' hearing aids for speech processing in noisy listening conditions. Gains in speech intelligibility could be predicted from gameplay accuracy and baseline inhibitory control. However, benefits did not persist in the absence of continuing practice. These studies employ stringent clinical standards to demonstrate that perceptual learning on a computerized audio game can transfer to "real-world" communication challenges.

INTRODUCTION

Sensorimotor skills can be acquired and refined throughout adulthood. This form of implicit learning is thought to depend, at least in part, on structural, neurochemical, and functional changes in sensory and motor regions of the adult cortex that emerge with practice on reinforced sensory or motor tasks [1]. Whether and how these plasticity mechanisms can be engaged by simple, computerized "brain-training" games to drive enhanced cognitive and perceptual abilities is a subject of intense debate [1, 2]. Perceptual training paradigms are typically psychophysical tests with behavioral feedback added at the end of each trial. These paradigms drive threshold improvements, though these gains generally do not transfer far beyond the training stimuli [3, 4]. For many perceptual training studies, the specificity of learning is a feature, not a bug, that can be used to infer the relative involvement of different brain regions as well as underlying plasticity mechanisms that enable and constrain performance [4–10]. But for more clinically oriented studies that set out with the goal of imparting a broad spectrum of enhanced perceptual abilities as a means to forestall the deleterious effects of aging or sensory impairment, the specificity of learning is a curse [11]. As an example, in the auditory modality, there is a strong motivation to improve communication abilities in older adults by boosting the intelligibility of target speech occurring in high levels of background noise. For the most part, this has been attempted by training hearing-impaired subjects to discriminate variations in low-level speech features using adaptations of psychophysical testing procedures. As with most any conventional perceptual training protocol, substantial improvements are noted on practice stimuli (~40%), but speech discrimination benefits are highly specific and show minimal transfer to untrained words [12–17], or even trained words presented in the context of untrained sentences [18].

Generalized gains in perceptual processing are routinely reported when training stimuli are instead packaged as games that require subjects to shift their focus of attention between multiple targets and devise fluid motor strategies for continuous, dynamic sensory challenges. A growing literature reports that relatively short periods of training with action video games drives enhanced visual processing across psychophysical tasks ranging from low-level feature detection to spatial attention

[19–21]. Importantly, perceptual gains through action game training can have therapeutic value. For example, amblyopic subjects show striking improvements in acuity after training with specific formats of action video games, even when their age is beyond the critical period for conventional rehabilitation therapies such as eye patching [22, 23]. Generalized perceptual improvements need not be limited to the visual modality. A history of musical training has been associated with a wide-range of enhanced auditory perceptual abilities, ranging from low-level feature discrimination to speech processing in noise [24–27], although, like action video games, the mechanisms driving these improvements are not fully understood [20, 28, 29]. Here we ask whether some of the challenges inherent to playing musical instruments or action video games could be packaged into a computerized audiomotor training interface to promote the generalized gains in speech processing that have proven elusive in prior auditory training studies [17].

In conventional perceptual training tasks, subjects react to stimuli presented in discrete trials. When playing video games or musical instruments, there are no trials. Instead, the player continuously monitors discrepancies between actual and predicted changes in sensory input generated through a closed loop (CL) between their motor actions and sensory feedback. Shifting the role of the subject from that of occasionally reacting to stimuli out of their control to continuously monitoring predicted changes in sensory feedback linked to their movement has far-ranging implications for the nature and form of neural processing leveraged to solve motor and perceptual discrimination challenges [30–33]. Action video games and musical training are also structured to encourage self-mastery and confidence for players with variable entry-level abilities by supporting incremental learning via precisely timed, frequent behavioral reinforcement [19]. Compared to the sparse, humdrum reinforcement provided in typical ''gamified'' versions of laboratory tasks, the high rates of emotionally salient reward and reinforcement prediction errors that are baked into action video games and musical training may more powerfully recruit sub-cerebral neuromodulatory centers that enable learning-related plasticity in adult sensory cortex [1, 34–37].

With these features in mind, we programmed an auditory training task that encouraged careful real-time monitoring of sensorimotor prediction errors and provided frequent behavioral reinforcement. We tested whether training on this CL audiomotor task was associated with generalized benefits in auditory perception and whether these gains, once learned, remained in the absence of ongoing practice. We performed our training in hearing-impaired older adults because they are often the target of marketing efforts for ''brain-training'' software and because neither existing training software nor their hearing aids offer reliable assistance with ''real-world'' communication signals, such as understanding speech in a crowded room [13, 38]. We carried out our study as a double-blind, randomized, placebo-controlled trial. To this end, we programmed a second auditory training interface that focused on improving memory capacity for words in spoken sentences, as auditory working memory (WM) is thought to be essential for speech recognition and has been a focus of prior training exercises [39–41]. As described below, subjects in both training groups reported similar expectations for improved speech processing, confirming that any placebo effects would be matched between training groups [42].

RESULTS

Unsupervised Auditory Learning in Older, Hearing-Impaired Subjects

We enrolled older adults ($\bar{x}$ = 70 years) living with mild to severe sensorineural hearing loss in a double-blind, randomized, placebo-controlled study (Figure 1A; Figures S1 and S2). We programmed an integrated suite of software applications that managed remote, unsupervised auditory training and psychophysical testing via a tablet computer [43]. Participants were randomized to the CL or WM group and asked to train for approximately 3.5 hr per week for 8 weeks while wearing their hearing aids (WM group: n = 11, $\tilde{x}$ = 31 hr total; CL group: n = 13, $\tilde{x}$ = 35 hr total; z = −1.3, p = 0.2, Wilcoxon rank-sum test). All subjects were long-term, bilateral hearing aid users (mean period of hearing aid use = 7 years).

The overarching objective for both tasks was to reconstruct jigsaw puzzles using a touchscreen interface on a tablet computer. Individual puzzle pieces were earned by solving auditory tasks. In the WM task, subjects were challenged to retain keywords from multi-talker spoken sentences in WM during a 3–16 s delay and then use the task interface to link together keywords from individual speakers to reassemble puzzles (Figure 1B, top; Figure S3). The difficulty level of the task adaptively changed such that advancing to later puzzle boards imposed higher memory loads, longer delay periods, and additional distractor elements (Figure 1B, middle and bottom). The CL audiomotor training task was modeled after earlier work on CL training tasks in animal [30–32, 44] and human [30] subjects. In brief, subjects discriminated subtle changes in continuous auditory feedback to trace the outline of the hidden puzzle piece (Figure 1C, top), place the puzzle piece in its correct position within the puzzle board (Figures S4A–S4C), and then rotate the piece into the correct orientation (Figures S4G–S4I). As with a musical instrument, movement of the stylus or fingertip on the touchscreen was converted into instantaneous auditory feedback that was used to update predictions about the current position relative to points of interest on the game board. Solving each CL game board required subjects to discriminate subtle variations in sound level, frequency, or modulation rate of tone pips or spectrotemporally modulated noise presented in a background of speech babble. We reasoned that continuously monitoring subtle differences between actual and predicted sensory feedback while suppressing complex, fluctuant background masking noise captured several of the cognitive challenges associated with processing speech in noise. As subjects became more proficient using CL auditory feedback to trace, place, and rotate invisible objects, the task was made more challenging by progressively increasing the level of background speech babble (Figure 1C, middle and bottom).

Although the cognitive and sensory demands of the WM and CL tasks were quite different, subjects rated both tasks similarly for immersion, overall difficulty, and expectation of benefits for improved speech perception (p > 0.56 for all comparisons; Figure S5). Subjects that trained in the WM task advanced to higher levels of the game, resulting in a significant increase in WM load

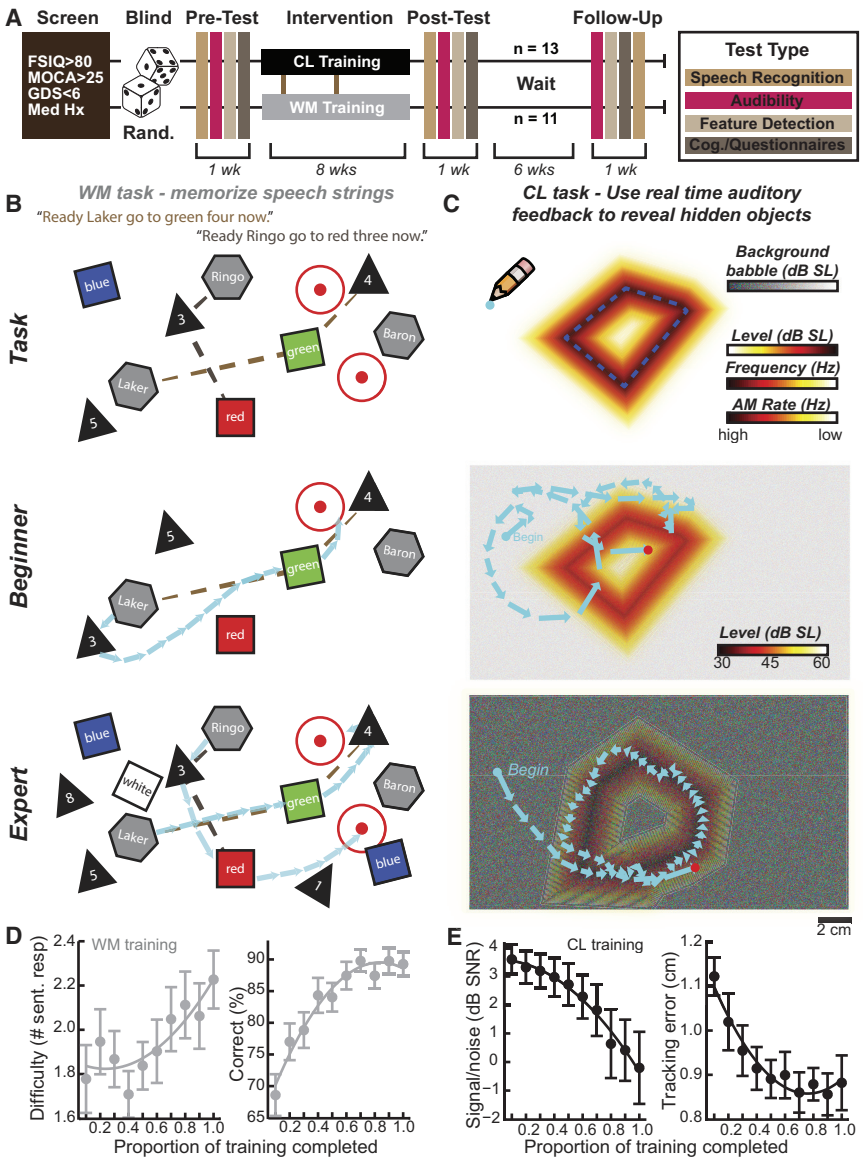

Figure 1. A Randomized Control Trial to Characterize Sound Perception before and after Training on WM or CL Audiomotor Discrimination Tasks

(A) Older adults who met eligibility requirements (brown box) were assigned to train on the working-memory (WM) or closed-loop (CL) audiomotor task through a randomized (Rand.) double-blind control trial. Subjects underwent testing with psychophysical measures and questionnaires before, during, just after, and 2 months after training (color coded by subject area).

(B) The WM game challenged subjects to maintain keywords from spoken sentences in WM and then linked together floating word elements from corresponding speech strings. Correct responses are indicated here by the colors of the spoken sentence (top). Blue arrows show finger movements on the touchscreen from early (middle) and late (bottom) in training that depict more accurate performance and increasing task complexity over time.

(C) The CL task challenged subjects to maneuver their virtual pencil within a sound gradient in an effort to maintain their position over the lowest sound level, highest frequency, or fastest amplitude modulation (AM) rate; see color scale bars, top). Target auditory features were presented in a background of continuous, distracting speech babble. Subjects learned to use real-time changes in sound level, frequency, or AM rate to more accurately trace the outline of more complex shapes that were "hidden" in progressively higher levels of speech babble noise. Blue arrows represent the distance and trajectory of the virtual pencil over 0.5 s sample periods for a single subject on day 8 (middle) and day 28 (bottom) of training.

(D) The WM task became more difficult over the course of training. Difficulty was defined as memory load, i.e., the number of sentences that required a response (left). WM performance improved over the course of training on a fixed "yardstick" condition that required subjects to respond to one of two sentences presented with three additional distractors on the screen (right).

(E) The CL tracing task became more difficult over training as the available SNR cues to identify the outline of the hidden shape decreased (left). Tracking error decreased on a CL "yardstick" condition over the course of training (right).

Circular symbols and error bars reflect the mean ± SEM. Overlying lines are quadratic fits to the group averaged data. See also Figures S1–S5.

experienced over the course of training (Figure 1D, left; $F = 2.2$, $p = 0.03$, repeated-measures ANOVA). Similarly, background speech babble levels increased as subjects advanced through higher levels of the CL game, resulting in a significant decrease in the signal-to-noise ratio (SNR) of target auditory cues over the course of training (Figure 1E, left; $F = 9.2$, $p = 3 \times 10^{-10}$, ANOVA). To assess the degree of learning in each task, we occasionally presented a "yardstick" task board to subjects over the course of training that offered a fixed degree of WM or auditory SNR challenge. Subjects in both training groups demonstrated a 20%–30% improvement in performance over the two-month training period in their respective tasks (WM, Figure 1D, right: $F = 14.2$, $p = 2 \times 10^{-13}$; CL, Figure 1E right: $F = 7.4$, $p = 2 \times 10^{-8}$; repeated-measures ANOVA). As a whole, both groups were tentatively optimistic that their training would improve their hearing, and about half believed that their hearing had improved during the study (Figure S5). Therefore, both the CL and WM tasks became more difficult as subjects advanced in their game play, subjects showed comparable levels of learning in both tasks, and subjects' game play experience and expectations for benefits through training (i.e., placebo effects) were matched [42].

Transfer of Learning to Speech-in-Noise Tasks after CL, Not WM, Training

The principal motivation for this work was to determine whether the benefits of auditory training transferred to "real-world" communication challenges, such as speech processing with

Table 1. Experimental Hypotheses

Means of Improvement	Which Training Group Will Show the Benefit?	Which Psychophysical Test(s) Will Reveal the Benefit?
Hypothesis 1: improved speech processing would arise from enhanced low-level auditory feature processing	CL, not WM	FM detection threshold; individual words, digits, or sentences in low-SNR conditions
Hypothesis 2: improved speech processing would arise from enhanced cognitive processing of linguistic cues in speech	WM, not CL	lower-context sentences in all SNR conditions; improved WM capacity
Hypothesis 3: improved speech processing would arise from enhanced stream segregation for target and background talkers	CL, not WM	all sentences or digit sequences in low-SNR conditions
Hypothesis 4: improved speech processing would arise from enhanced motivation and confidence for speech testing (placebo effect)	WM and CL	all speech-in-noise tests

competing talkers in the background. Prior work has suggested poor speech-in-noise intelligibility may reflect a combination of deficits in low-level feature discrimination, selective attention, or other cognitive abilities [45–47]. Therefore, it was conceivable that improved proficiency in either task could be associated with improved speech recognition accuracy. We investigated this possibility by having subjects complete a corpus of speech recognition tasks that varied according to (1) SNR, (2) duration (e.g., individual words in noise versus sentences in noise), and (3) linguistic cues, ranging from sentences with higher context (e.g., "The janitor swept the floor"), lower context (e.g., "Dimes showered down from all sides"), or no context (e.g., single words or strings of random digits). Further to this, we also asked subjects to complete psychophysical tests that measured temporal fine structure detection thresholds, inhibitory control, and WM capacity. We hypothesized any transfer of learning from either of the training tasks to the "real-world" challenge of speech processing in noise could arise from four different processes, each carrying a unique prediction as to what combination of training task and testing conditions would reveal improvements (Table 1).

Our results strongly supported the third hypothesis described in Table 1; we found that training in the CL task, but not the WM task, led to generalized improvements in speech processing for all types of low-SNR stimuli that built up over time (group × session × measurement type interaction: $F = 3.14$, $p < 0.005$, mixed-model ANOVA; see Data S1 for a detailed presentation of statistical outcomes). We focused our analysis on challenging listening conditions that approximated a crowded restaurant [48], where speech intelligibility was greatly reduced, but not impossible (Figures 2A–2C, left, red vertical lines). We observed that recognition of isolated words at difficult SNRs (Figure 2A, right) did not improve secondary to either training approach (group × session interaction: $F = 0.14$, $p = 0.71$; main effects for session in both CL and WM training groups: $F < 2.68$, $p > 0.1$; repeated-measures ANOVA). However, when speech processing was measured with sentences rather than isolated words, we noted an ~25% increase in accuracy within the low-SNR "intelligibility cliff" for the CL, but not the WM, training group. This benefit was noted both for low-context (28.5%; Figure 2B, black) and high-context (21.7%; Figure 2C, black) sentences (training group × session interactions: $F > 5.46$ 0.14, $p < 0.03$; main effects for session: $F > 8.67$, $p < 0.02$ for both lower- and higher-context sentences; repeated-measures ANOVA). By contrast, WM training was not associated with any significant change in the percentage of correctly identified words in lower-context (−4.7%; Figure 2B, gray) or higher-context (4.04%; Figure 2C, gray) sentences (main effects for session in both lower- and higher-context sentences: $F < 0.76$, $p > 0.4$, repeated-measures ANOVA). In summary, we observed a selective transfer of learning to untrained sentence—but not word—recognition tasks only in subjects who trained on the CL task (Figure 2D).

As further evidence that CL training benefits did not reflect an enhanced ability to make use of semantic or syntactic cues, we also asked subjects to identify four random digits spoken by a target talker while two other talkers that differed in fundamental frequency (F_0) simultaneously produced competing digit streams (Figure 2E). Like sentence tests, this task required attentional selection of a target speech stream that built up over longer timescales, but in contrast to sentence tests, there were no linguistic cues. For the sake of direct comparison with the speech tests described above, we broke down their response according to whether the first digit was correctly identified (akin to words in noise) or the entire sequence of digits was correctly identified (akin to sentences in noise) and expressed the effect of training as a change index that was bounded from −1 to +1, where negative values indicated reduced speech performance after training and positive values indicated improved performance. As per the selective benefit of CL training on sentence intelligibility, the difference in training group emerged when accuracy was assessed across the complete sequence of digits, but not when it was measured for the first digit (main effect for training group on first digit/word change indices: $F < 0.01$, $p > 0.97$; main effect for training group on full digit sequence/sentence change indices: $F > 4.8$, $p < 0.05$; training group effect across all test types: $F = 4.9$, $p = 0.12$; multivariate and univariate ANOVA; Figures 2D and 2F). Moreover, digit streaming improvements were significantly correlated with sentence-in-noise recognition benefits ($R = 0.72$, $p = 0.01$).

Improved Audio Tracing Skill Predicts the Degree of Enhanced Speech Processing

We programmed the auditory CL training task to reflect some of the challenges that might be encountered when playing a

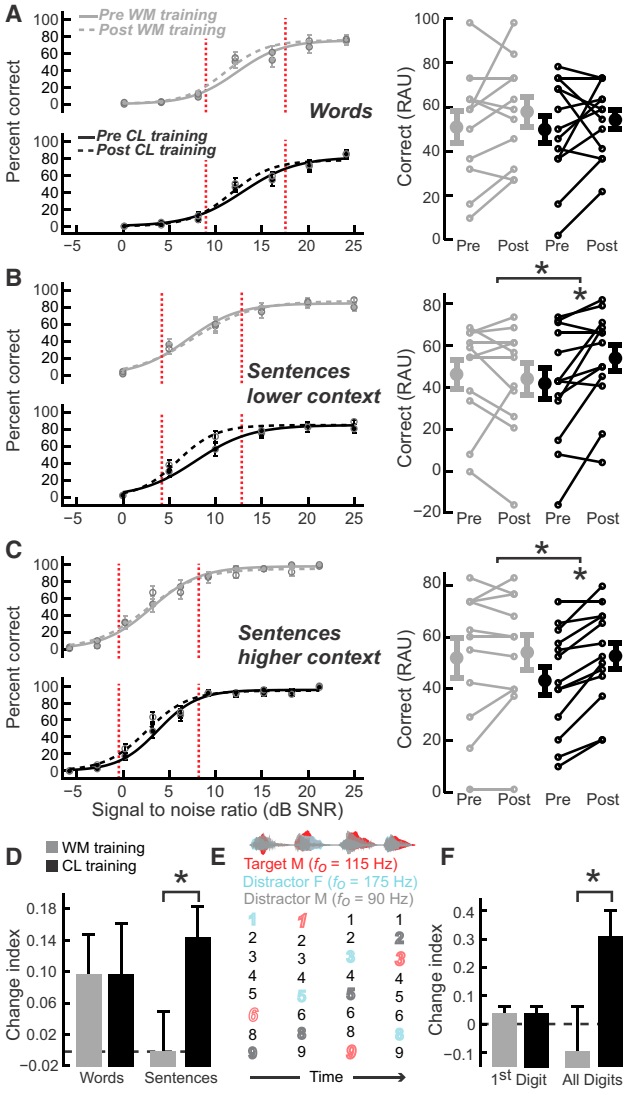

Figure 2. CL Training, but Not WM Training, Was Associated with Generalized Improvements in Sentence Comprehension in High Levels of Background Noise

(A–C) Recognition of monosyllabic words (A), lower-context sentences (B), or higher-context sentences (C) assessed before (solid lines) and after (dashed lines) training with the WM (gray) or CL (black) training as a function of background noise levels. Left: accuracy was measured at each SNR and average performance data were fit with a logistic function using constrained maximum-likelihood estimation. Recognition performance declined steeply within a restricted set of SNRs (vertical red broken lines). Right: smaller lines reflect individual subject pre- and post-test scores for SNRs that fall within the vertical red lines, shown at left. Larger circles and error bars represent the mean ± SEM for each training group.

(D) Summary plot of primary outcome measures expressed as a change index, $(Post\ score(\%) - Pre\ score(\%)/Post\ score(\%) + Pre\ score(\%))$, where a value of zero indicates no change after training.

(E) Schematic of digit streaming task. Male target speaker waveform and target digits (red) are depicted alongside distractor male (gray) and female (cyan) speech waveforms and spoken digits. f_o, fundamental frequency (voice pitch).

(F) Summary of digit task improvements when the first digit is scored in isolation, similar to a word test, or the whole stream of digits is scored, similar to sentence comprehension tests. Bar plots reflect the mean ± SEM.

Asterisks indicate statistically significant differences between training groups at $p < 0.05$. See also Tables S1 and S2.

musical instrument. As a violinist would produce a target pitch by adjusting finger position on the neck of the instrument according to real-time auditory feedback, the CL task also challenged our subjects to use subtle changes in the frequency, level, and modulation rate of sounds to update their finger position until the target sound was produced. As our violinist would focus on the sound of their instrument while suppressing the din of the surrounding orchestra, subjects in the CL task were also challenged to suppress the distraction of increasingly loud background speech babble and direct their attention instead to feedback cues linked to their movement. In this regard, some of the essential perceptual challenges faced by an orchestra violinist and an elderly, hearing-impaired person engaged in the CL audiomotor task are the same: to segregate target and distractor streams and then utilize auditory error predictions to update forward motor models (Figure 3A). If improved sentence processing in background noise was linked to the demands of the CL training task, we expected that subjects who showed the largest improvements in speech processing would be the ones who also demonstrated the greatest improvement in using instantaneous audio feedback to trace the outline of hidden shapes (Figures 3B and 3C). Indeed, we observed that the reduction in auditory tracing error in the CL task was significantly correlated with improved speech reception thresholds (Figure 3D; $R = 0.60$, $p = 0.03$).

Audiomotor Training Enhances Speech Intelligibility in Levels of Background Noise at which Hearing Aids Provide Little Benefit, but the Effects Do Not Last

Difficulties processing speech in background noise is nearly unavoidable in older adults [45, 49, 50]. Any benefit that our subjects received from their hearing aids tapered off as background noise levels reached what would be encountered in a typical social environment (Figures 4A and 4B). To compare the benefit provided by participants' hearing aids alone versus the benefit provided by their hearing aids plus CL training, we calculated the difference in speech recognition accuracy under aided or unaided listening conditions as well as before and after training. We found that 8 weeks of CL training provided approximately three times the benefit of the hearing aids that subjects had been wearing for approximately 7 years, but only at low SNRs (Figure 4B; low SNR [0–6 dB]: $t = -5.19$, $p = 3 \times 10^{-4}$; high SNR [9–21 dB]: $t = 0.87$, $p = 0.4$; unpaired t test). In this sense, CL auditory training is a useful adjuvant for hearing aids, in that speech processing was specifically enhanced in the SNR listening environments where their hearing aids offer little benefit but that represent a chief complaint of patients [50].

In clinical terms, the efficacy of a hearing aid or other prosthetic device is judged to be significant for an individual patient if it provides benefit that substantially exceeds the inherent variability of the outcome measurement (defined statistically as the critical difference). We evaluated the proportion of participants in each treatment group who experienced a significant improvement in speech reception thresholds secondary to intervention. We found that 8 weeks of CL training imparted a significant speech processing benefit to 62% of subjects, whereas only 9% of subjects trained on the WM task met this criterion (Figure 4C). To probe the persistence of training benefits, we performed a third set of follow-up tests after approximately 7 weeks

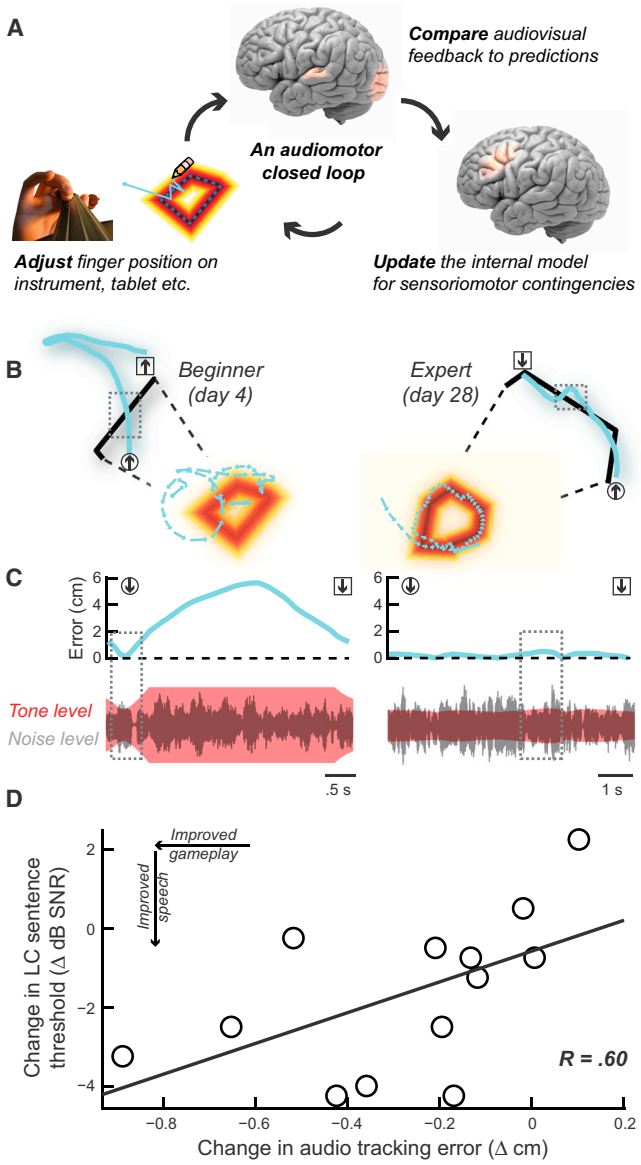

Figure 3. Improvements in Speech Processing after CL Training Can Be Predicted from Game Play Performance

(A) Cartoon illustrating how a CL between continuous auditory discrimination and adjustments in finger position create sensory prediction errors that can be used to produce target sounds. A similar process underlies learning to produce a target sound on the violin or trace the hidden auditory outline in the CL tablet task.

(B)Tracing segments from early and late CL game boards illustrate improvements in audio tracking over the course of training. Tracing error, defined as the instantaneous Euclidean distance measured between each point in the virtual pencil trace (cyan) and the shape border (black), is plotted over an ~3–7 s tracing period. The beginning and end of the tracing segment are represented by an arrow encased by a circle and square, respectively.

(C) For the SNR tracing task, the distance between the virtual pencil and the hidden border is instantly translated into the SNR of the tone in noise. Note that the tone level gradient saturates outside of a 2 cm area surrounding the shape outline (C, left bottom). The gray rectangles in (A) and (B) focus on a specific portion of the trace, highlighting the relatively subtle changes in the signal envelope used to refine finger movement trajectory in this well-trained subject (compare C, left and right). Signal magnitude (bottom) is plotted on a logarithmic scale for ease of visualization.

(D) Reductions in audio tracing error over the course of CL training are associated with improved sentence recognition threshold (lower indicates better performance). R, Pearson's correlation coefficient.

without any additional training. We computed the statistical effect sizes for CL relative to WM treatment at the post-training and 2-month follow-up tests. While we observed large effect sizes for both sentence-in-noise tests as well as the digits streaming task at the conclusion of training (lower-context sentences = 0.88, higher-context sentences = 0.85, digit stream = 0.94, Hedges' g), the difference did not persist when tested in the absence of further intervention 2 months later (Figure 4D; Hedges' $g < 0.14$). These results suggest that, at least with this sensory-impaired elderly population, generalized gains in speech perception through CL training may require some degree of "topping off" through continuing practice.

Inhibitory Control Does Not Improve with Training and does Not Predict Speech Recognition Ability, but Does Predict which Subjects Will Benefit the Most from CL Training

As a final analysis, we determined whether any baseline psychophysical measurements could predict which subjects benefit the most from CL training or were most likely to have poor sentence-in-noise processing. We reasoned that subjects who already demonstrated strong native abilities to suppress task-irrelevant information would stand to benefit the most from a training task that emphasized fine-grained acoustic discrimination while ignoring background speech babble. We measured inhibitory control by administering visual and auditory versions of the Stroop task before and after training. In the visual version of the task, the reaction time to correctly report the font color of a neutral word, "legal," was compared to the words "blue" or "red" when the font color either matched or did not match the corresponding word (congruent and incongruent, respectively). In an auditory variant of this task, the time to correctly report the voice pitch of a neutral word, "day," was compared to the words "high" or "low" (Figure 5A).

Overall, we observed a robust Stroop effect in the auditory modality, in that reaction time was accelerated relative to the neutral condition when the speaker's pitch matched the presented word but was substantially slower when the voice pitch was mismatched (Figure 5B). As expected, subjects with the best baseline inhibitory control (i.e., the smallest effect of distractor congruence on reaction times) demonstrated the greatest CL training benefits from improved processing of sentences in noise (combined Stroop: R = 0.66, p = 0.01, Pearson's correlation coefficient; Figures 5C and S6C). The baseline congruence cost was not correlated with baseline sentence recognition threshold (audio: R = 0.10, p = 0.61; visual: R = 0.17, p = 0.78; Pearson's correlation coefficient corrected for multiple comparisons with the Holm-Bonferroni method). Likewise, we did not find that normalized performance on either version of the Stroop task improved secondary to task intervention (training group × session interactions, audio: F = 1.04, p = 0.32; visual, F = 0.17, p = 0.68; repeated-measures ANOVA; Figures 5D and 5E), though overall reaction times decreased significantly (~50–100 ms) for both the CL and WM

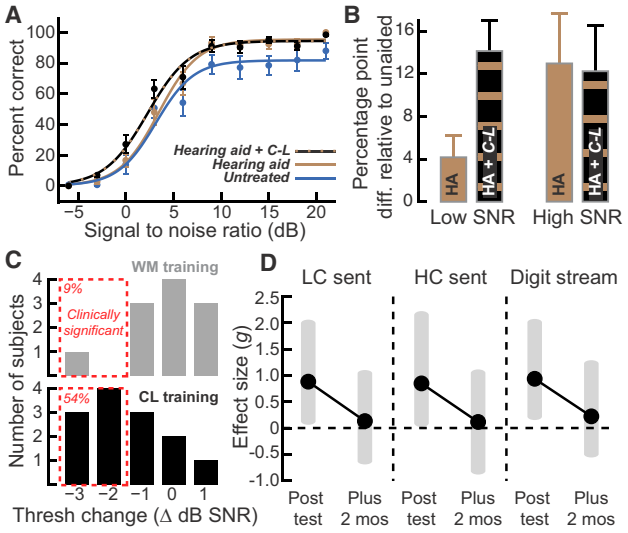

Figure 4. Audiomotor Training Temporarily Enhances Speech Perception at Background Noise Levels at which Hearing Aids Provide Little Benefit

(A) Recognition of higher-context sentences was measured at SNRs that ranged from poor to favorable without hearing aids (blue), while using their hearing aids before CL training (gold), and when using a hearing aid after CL audiomotor training (black and gold).

(B) The percentage point difference in speech accuracy provided by a hearing aid and a hearing aid plus CL training under low-SNR (0–6 dB) and high-SNR (9–21 dB) conditions. HA, hearing aid; CL, closed-loop audio training. Error bars represent the SEM.

(C) Histograms of sentence recognition threshold changes (negative values indicate better performance) after WM and CL training (gray and black, respectively). Threshold changes of 1.3 dB SNR or greater are considered "clinically significant" (red dashed lines).

(D) CL gaming effect sizes and 95% confidence intervals (gray shaded areas) for sentence (LC, lower context; HC, higher context) and digits streaming measurements after CL training and at the 2-month follow-up visit.

groups (Figure S6B). Thus, measures of inhibitory control could predict the subjects who benefited the most from training but did not predict baseline speech-in-noise thresholds and did not change as a result of training.

Frequency Discrimination Thresholds and WM Capacity Predict Speech-in-Noise Recognition Accuracy, but Do Not Change after Training and Do Not Predict Learning Transfer

Measurements that assess low-level detection thresholds for changes in temporal fine structure as well as high-level cognitive functions, such as WM, have been shown to predict sentence-recognition-in-noise performance for individuals with and without hearing loss [45–47]. Logically, a change in speech-in-noise intelligibility through training could be accompanied by a commensurate change in these measures as well. We measured spectrotemporal processing ability by assessing sensitivity to periodic fluctuations in the frequency of a pure tone (frequency modulation [FM]), a task that has been shown to reflect accurate encoding of temporal fine structure cues [51]. We also measured WM capacity by administering the Letter-Number Sequencing test, which involves the repetition and ordering of alphanumeric strings of increasing length [21]. Consistent with previous reports, we found that both basic spectrotemporal processing abilities and auditory WM capacity were significantly correlated with baseline sentence-in-noise recognition (FM: R = 0.44, p = 0.04, Figure 5F; WM: R = 0.45, p = 0.04, Figure 5G; Pearson's correlation coefficient corrected for multiple comparisons with the Holm-Bonferroni method). However, neither low-level FM detection nor cognitive WM scores were significantly changed in either training group (training group × session interaction, FM: F = 0.31, p = 0.58, Figure 5H; WM: F = 0.18, p = 0.68, Figure 5I; repeated-measures ANOVA).

Taken as a whole, these correlations suggest that the abilities that best predict native speech-in-noise processing are independent of the abilities that predict the amount of training-related improvements in speech processing. This suggests some degree of independence between brain systems that support learning versus those that support performance, at least in the baseline pre-training condition. That WM improvements gained through the training interface did not generalize to other tests of WM (Figure 5I, gray lines) confirms the central critique leveled against many types of cognitive "brain-training" exercises [2, 29]. At the same time, the fact that generalized gains in speech processing were observed with CL training underscores the hazards of throwing the baby out with the bathwater; at least in the context of perceptual enhancements imparted through CL sensorimotor training tasks, a spectrum of clinically meaningful improvements in aural speech intelligibility can be imparted through computerized training in sensory-impaired older adults.

DISCUSSION

We programmed a suite of self-administered tablet-based auditory training and psychoacoustic testing software that could be used from home by older adults with sensorineural hearing loss. Prior efforts to improve speech outcomes in older, hearing-impaired subjects through speech discrimination training have described strong on-task learning accompanied by changes in underlying neural processing but only modest perceptual transfer to untrained speech materials [10, 12, 14–16, 18]. The weak transfer of learning combined with a reliance on poorly controlled study designs have raised doubts about the utility of auditory training as a means to improve speech outcomes in persons with hearing impairment [17]. Our work was motivated by two goals: (1) to implement a placebo-controlled study design and (2) to use a different format for computerized auditory training that was more consistent with neuroscience-based principles for driving positive brain plasticity than "gamified" adaptations of audiological or psychoacoustic tests.

Designing a "Sugar Pill" to Control for Subject Expectations and Motivation in Perceptual Training Studies

In pharmaceutical clinical trials, subjects assigned to the control group are administered sugar pills that are indistinguishable from the experimental pills. Because both the subject and the study team are blind to the group assignment, subjects in both groups have equivalent expectations for benefit and equivalent motivation to adhere to the strictures of the trial. This defines the "gold standard" for a control, in that the experimental treatment is the

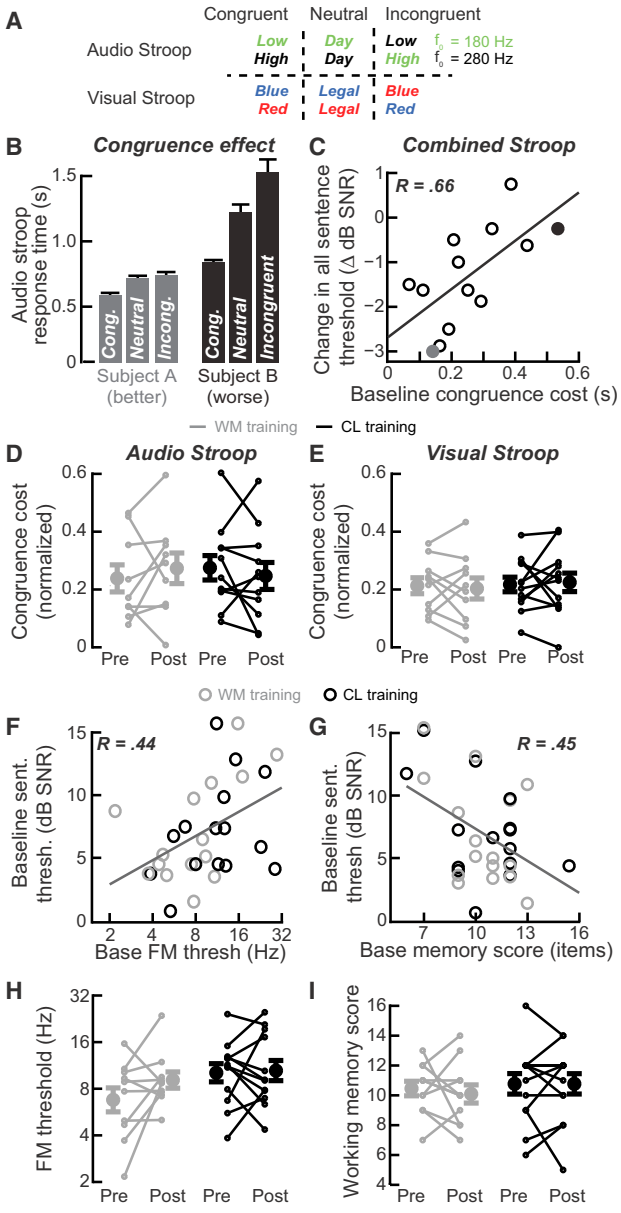

Figure 5. Other Psychophysical and Cognitive Measures Can Predict Training Benefit or Baseline Speech Recognition Scores, but None Change after Training

(A) Schematic of the audio and visual Stroop test conditions. F_0, fundamental frequency.

(B) Average reaction times for each congruency condition in the audio Stroop task from two subjects: one with low congruence cost (left, gray) and the other with high congruence cost (right, black). Error bars represent the SEM.

(C) The effect of congruence on response time is computed as the congruence cost $(Incongruent\ RT(s) - Congruent\ RT(s)/Nuetral\ RT(s))$. The mean baseline congruence cost for auditory and visual versions of the Stroop task predicts the degree of improvement in sentence processing after CL training (data from individual subjects in B are plotted with the corresponding black and gray shading). Improved speech processing is defined as the change in the SNR at which subjects correctly perceive 50% of the words. R, Pearson's correlation coefficient.

(D and E) Inhibitory control, in the form of auditory (D) or visual (E) congruence cost, does not change after WM or CL training.

(F–I) Frequency modulation detection thresholds (F) and WM scores (G) are correlated with baseline sentence-recognition-in-noise abilities but do not change as a result of training (H and I; gray, memory training; black, CL training).

See also Figure S6.

only factor that can account for systematic differences between groups. Auditory training studies in hearing-impaired subjects have generally fallen short of this standard by relying on a passive, "no contact" control group or no control group at all. In these cases, interpreting the basis for change before and after training is difficult, as subjects' expectations for benefit and motivation to participate would also be expected to differ systematically between the treatment and control groups. Several recent studies have used an "active control" group, where subjects are assigned an activity during the training period [13, 15]. Even active controls fall short of the mark because there is no explicit demonstration that the control group is matched for factors that are not directly relevant for driving speech processing improvements [42]. An approach that more closely approximates an auditory training "sugar pill" would be to train a control group on a complex speech discrimination task that would not be expected to provide any generalized benefit for processing speech in noise but would leave subjects with an equivalent expectation for improved hearing.

In our study, subjects were randomly assigned to the CL or WM training group, where both the subjects and study staff were blind to the treatment group assignment. Both the WM and CL tasks used adaptive tracking to adjust the difficulty in their respective task as performance improved (Figure 1D). Subjects in both groups showed comparable evidence of in-task learning (Figure 1E) and equivalent expectations for improved hearing abilities by the end of training (Figure S5A). Additionally, both groups reported equivalent immersion in their training tasks (Figure S5B), and belief that their course of therapy was helping (Figures S5C and S5D). By demonstrating that the WM control group was matched for adaptive training challenge, in-task learning, and expectations, we met the high standard of a placebo control for a sensory training intervention. By controlling for factors unrelated to the treatment effect, we can more efficiently home in on the distinguishing features of the CL training task that may have supported the transfer of learning to unpracticed speech materials.

CL Audiomotor Training Improves Performance in "Real-World" Listening Challenges

We focused on training older hearing aid users, a demographic that often struggles to follow conversations in noise and for whom neither hearing aids nor currently available auditory training software offer reliable assistance [13, 38]. Our training interface empowered subjects to control the frequency spectrum, envelope modulation rate, and level of dynamic sounds while closely monitoring errors between actual and predicted changes in sensory feedback. As performance improved, target sound features were embedded in increasingly high levels of background speech babble. This task design encouraged subjects to focus their attention on subtle variations in low-level acoustic features while segregating and suppressing the distracting influence of background speech. The target stimuli in

the CL training task were not speech, but rather dynamic tones, "tone clouds," or spectrotemporally modulated "ripple" noise that were tailored to each subject's audiometric and comfortable loudness thresholds. We used the native microphone and camera hardware in the tablet computer to monitor speaker positioning and ambient sound levels during training. The software was programmed to randomly assign elderly hearing-impaired subjects to training groups, encrypt their identity, guide them through self-directed psychophysical testing, and administer questionnaires while maintaining double-blind testing conditions. Collectively, these efforts demonstrate that home-based auditory training in sensory impaired older adults can be performed with placebo-controlled study designs and complex, engaging training interfaces.

We found that CL training provided a useful adjuvant for our subjects' hearing aids by enhancing speech intelligibility at low SNRs, where their devices offered little benefit. Numerically, the gains in speech recognition amounted to ~25% more words correctly identified in a given sentence, or a 1.5 dB SNR reduction in speech-in-noise recognition threshold. Importantly, these gains occurred within the steep "intelligibility cliff" of SNRs (vertical red lines in Figure 2), where even a small change in the number of correctly understood words can have a disproportionate effect on overall comprehension and communication experience. The effect size for speech outcome measures was fairly large (Hedges' g was 0.85–0.94 across speech tests; Figure 4D), yet training benefits were not observed for low-level acoustic discrimination thresholds or for cognitive measures of WM. The particular pattern of improvements on sentence-level speech processing without any transfer to either low-level sound processing or higher cognitive control or memory processes suggested that the CL training may have improved our subjects' ability to progressively focus the spotlight of auditory attention away from distractor sources and onto target speakers—a central requirement for maintaining a conversation in social environments [47, 52].

Considerations for Designing an Auditory Training Task

Training subjects on a restricted set of near-threshold stimuli using "gamified" versions of laboratory psychophysical tests is an ideal approach to prevent off-task, broadly generalized learning [8, 11]. For many basic science researchers, the specificity of learning is a necessary feature of their experiments. Work in these disciplines have embraced the "curse of specificity" by training animal models and human subjects on various derivations of psychoacoustic tasks to gain deeper insight into the biological and psychological factors that constrain the transfer and consolidation of learning [3–10]. Given that learning specificity is effectively baked into these tasks, it should come as no surprise that well-designed randomized control trials using derivations of perceptual training tasks find minimal evidence for a transfer of learning to untrained speech materials in hearing-impaired subjects [13, 17]. In the visual training literature, clinically oriented rehabilitation studies are increasingly turning toward tasks that promote a desire to learn by allowing the subject to navigate engaging, multi-layered training challenges built from diverse underlying stimulus sets [11, 19, 22, 23]. For the most part, the tasks used in the auditory training literature possess few, if any, of these qualities. These limitations could be overcome with more engaging computerized auditory rehabilitation strategies that emphasize a combination of executive control and low-level auditory feature discrimination, as also noted in a recent review of the auditory training literature [53].

Shortcomings and Directions for Further Improvement

With that said, there is no reason to believe that the design and implementation of the CL task described here represents an optimal solution to enhance speech-in-noise perception. Although the magnitude of improvements on untrained speech materials was fairly large (Hedges' $g > 0.8$), all benefits regressed to baseline within 2 months without additional training. On-task training effects are often retained for several months in younger [6] and older adult subjects [21], but generalized, off-task perceptual training benefits in sensory-impaired older adults have not been described, and therefore the expectations for the persistence of generalized learning are unknown. Regardless, without carefully probing the decay of learning with and without "topping-off" sessions in a larger, more diverse sample of subjects, it is unclear whether this CL auditory training task has any practical therapeutic value for persons with hearing impairment. To the contrary, it seems probable that larger, practically useful and more persistent training effects would be possible through iterative improvements in the design of the training interface. Sifting through the variance in our sample of subjects trained in the CL task revealed that the largest improvements in speech processing were observed in subjects with strong baseline inhibitory control (Figure 5A) who learned to modulate the speed and accuracy of their tracing to home in on subtle fluctuations in low-SNR acoustic cues (Figure 3D). This identifies a path for further improvements in the training interface to encourage successful gameplay strategies matched to native cognitive strengths.

Ultimately, the degree, generalization, and persistence of enhanced speech perception are behavioral proxies for the complex set of underlying physiological, chemical, and structural changes in the brain. In this regard, it isn't the training task, per se, that matters; the training interface only provides the means to non-invasively modulate signal processing in sensory, motor, and executive networks while engaging neuromodulatory centers that enable plasticity in the adult brain. The CL audiomotor training interface described here was designed to provide reward prediction errors distributed over a variety of timescales, a key feature for driving dopaminergic networks in the forebrain [37] and midbrain [34, 35] as well as cholinergic and non-cholinergic centers in the basal forebrain [34, 36]. We adapted our task from animal training studies that demonstrated large-scale improvements in auditory cortex coding of temporal envelope fluctuations [32] and low-SNR stimuli [30, 31, 54] after learning on CL audiomotor tasks. Looking forward, one can imagine a reverse-engineering approach to develop training tasks based on concurrent behavioral and neuroimaging measurements, wherein task design is not guided by teleological categories ("complex," "engaging," "cognitively challenging," etc.), but rather according to how well task elements recruit activity from targeted brain areas or promote the stabilization of a desired network state [55]. The ongoing development of principled, neuroscience-based approaches for perceptual training holds

enormous potential as a tool to rehabilitate and expand human perceptual and communication capabilities, whether on its own or in combination with assistive devices or pharmacological therapies [56, 57].

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subject Recruitment
- METHOD DETAILS
  - Subject enrollment and training procedures
  - Psychophysical tasks to assess transfer of learning
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and one table and can be found with this article online at https://doi.org/10.1016/j.cub.2017.09.014.

AUTHOR CONTRIBUTIONS

J.P.W., K.E.H., and D.B.P. designed the experiments. K.E.H. programmed the training and testing software. J.P.W. and J.M.S. performed the behavioral experiments. J.P.W. analyzed the data. J.P.W. and D.B.P. wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by NIH P50 DC015857 and a research grant from One Fund Boston (to D.B.P.). We thank Microsoft for the generous donation of Surface tablets.

Received: May 24, 2017
Revised: August 20, 2017
Accepted: September 11, 2017
Published: October 19, 2017

REFERENCES

1. Merzenich, M.M., Van Vleet, T.M., and Nahum, M. (2014). Brain plasticity-based therapeutics. Front. Hum. Neurosci. *8*, 385.
2. Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., and Ballard, C.G. (2010). Putting brain training to the test. Nature *465*, 775–778.
3. Fiorentini, A., and Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. Vision Res. *21*, 1149–1158.
4. Wright, B.A., Buonomano, D.V., Mahncke, H.W., and Merzenich, M.M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. J. Neurosci. *17*, 3956–3963.
5. Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron *36*, 791–804.
6. Mossbridge, J.A., Fitzgerald, M.B., O'Connor, E.S., and Wright, B.A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. J. Neurosci. *26*, 12708–12716.
7. Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. *26*, 4970–4982.
8. Hung, S.-C., and Seitz, A.R. (2014). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. J. Neurosci. *34*, 8423–8431.
9. Hawkey, D.J.C., Amitay, S., and Moore, D.R. (2004). Early and rapid perceptual learning. Nat. Neurosci. *7*, 1055–1056.
10. Tremblay, K.L., and Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. J. Speech Lang. Hear. Res. *45*, 564–572.
11. Deveau, J., and Seitz, A.R. (2014). Applying perceptual learning to achieve practical changes in vision. Front. Psychol. *5*, 1166.
12. Burk, M.H., and Humes, L.E. (2008). Effects of long-term training on aided speech-recognition performance in noise in older adults. J. Speech Lang. Hear. Res. *51*, 759–771.
13. Saunders, G.H., Smith, S.L., Chisolm, T.H., Frederick, M.T., McArdle, R.A., and Wilson, R.H. (2016). A randomized control trial: supplementing hearing aid use with listening and communication enhancement (LACE) auditory training. Ear Hear. *37*, 381–396.
14. Kuchinsky, S.E., Ahlstrom, J.B., Cute, S.L., Humes, L.E., Dubno, J.R., and Eckert, M.A. (2014). Speech-perception training for older adults with hearing loss impacts word recognition and effort. Psychophysiology *51*, 1046–1057.
15. Anderson, S., White-Schwoch, T., Choi, H.J., and Kraus, N. (2013). Training changes processing of speech cues in older adults with hearing loss. Front. Syst. Neurosci. *7*, 97.
16. Ferguson, M.A., Henshaw, H., Clark, D.P., and Moore, D.R. (2014). Benefits of phoneme discrimination training in a randomized controlled trial of 50- to 74-year-olds with mild hearing loss. Ear Hear. *35*, e110–e121.
17. Henshaw, H., and Ferguson, M.A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. PLoS ONE *8*, e62836.
18. Burk, M.H., Humes, L.E., Amos, N.E., and Strauser, L.E. (2006). Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. Ear Hear. *27*, 263–278.
19. Shawn Green, C., and Bavelier, D. (2015). Action video game training for cognitive enhancement. Curr. Opin. Behav. Sci. *4*, 103–108.
20. Green, C.S., and Bavelier, D. (2012). Learning, attentional control, and action video games. Curr. Biol. *22*, R197–R206.
21. Anguera, J.A., Boccanfuso, J., Rintoul, J.L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., and Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. Nature *501*, 97–101.
22. Li, R.W., Ngo, C., Nguyen, J., and Levi, D.M. (2011). Video-game play induces plasticity in the visual system of adults with amblyopia. PLoS Biol. *9*, e1001135.
23. Vedamurthy, I., Nahum, M., Bavelier, D., and Levi, D.M. (2015). Mechanisms of recovery of visual function in adult amblyopia through a tailored action video game. Sci. Rep. *5*, 8482.
24. Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. Ear Hear. *30*, 653–661.
25. Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A.J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. *219*, 36–47.
26. Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V., Kidd, G., Jr., and Patel, A.D. (2015). Musical training, individual differences and the cocktail party problem. Sci. Rep. *5*, 11628.
27. Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. Nat. Rev. Neurosci. *11*, 599–605.
28. Coffey, E.B.J., Mogilever, N.B., and Zatorre, R.J. (2017). Speech-in-noise perception in musicians: a review. Hear. Res. *352*, 49–69.
29. Bisoglio, J., Michaels, T.I., Mervis, J.E., and Ashinoff, B.K. (2014). Cognitive enhancement through action video game training: great expectations require greater evidence. Front. Psychol. *5*, 136.

30. Whitton, J.P., Hancock, K.E., and Polley, D.B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. Proc. Natl. Acad. Sci. USA *111*, E2606–E2615.
31. Polley, D.B., Heiser, M.A., Blake, D.T., Schreiner, C.E., and Merzenich, M.M. (2004). Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex. Proc. Natl. Acad. Sci. USA *101*, 16351–16356.
32. Bao, S., Chang, E.F., Woods, J., and Merzenich, M.M. (2004). Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. Nat. Neurosci. *7*, 974–981.
33. Wolpert, D.M., Diedrichsen, J., and Flanagan, J.R. (2011). Principles of sensorimotor learning. Nat. Rev. Neurosci. *12*, 739–751.
34. Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E.M., and Stephan, K.E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron *80*, 519–530.
35. Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., and Janak, P.H. (2013). A causal link between prediction errors, dopamine neurons and learning. Nat. Neurosci. *16*, 966–973.
36. Lin, S.-C., and Nicolelis, M.A.L. (2008). Neuronal ensemble bursting in the basal forebrain encodes salience irrespective of valence. Neuron *59*, 138–149.
37. Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E.M., and Graybiel, A.M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature *500*, 575–579.
38. Bentler, R.A., and Duve, M.R. (2000). Comparison of hearing aids over the 20th century. Ear Hear. *21*, 625–639.
39. Wayne, R.V., Hamilton, C., Jones Huyck, J., and Johnsrude, I.S. (2016). Working memory training and speech in noise comprehension in older adults. Front. Aging Neurosci. *8*, 49.
40. Kronenberger, W.G., Pisoni, D.B., Henning, S.C., Colson, B.G., and Hazzard, L.M. (2011). Working memory training for children with cochlear implants: a pilot study. J. Speech Lang. Hear. Res. *54*, 1182–1196.
41. Zhang, Y.X., Moore, D.R., Guiraud, J., Molloy, K., Yan, T.-T., and Amitay, S. (2016). Auditory discrimination learning: role of working memory. PLoS ONE *11*, e0147320.
42. Boot, W.R., Simons, D.J., Stothart, C., and Stutts, C. (2013). The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. Perspect. Psychol. Sci. *8*, 445–454.
43. Whitton, J.P., Hancock, K.E., Shannon, J.M., and Polley, D.B. (2016). Validation of a self-administered audiometry application: an equivalence study. Laryngoscope *126*, 2382–2388.
44. Williamson, R.S., Hancock, K.E., Shinn-Cunningham, B.G., and Polley, D.B. (2015). Locomotion and task demands differentially modulate thalamic audiovisual processing during active search. Curr. Biol. *25*, 1885–1891.
45. Füllgrabe, C., Moore, B.C.J., and Stone, M.A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. Front. Aging Neurosci. *6*, 347.
46. Mehraei, G., Gallun, F.J., Leek, M.R., and Bernstein, J.G.W. (2014). Spectrotemporal modulation sensitivity for hearing-impaired listeners: dependence on carrier center frequency and the relationship to speech intelligibility. J. Acoust. Soc. Am. *136*, 301–316.
47. Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B.G. (2011). Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. Proc. Natl. Acad. Sci. USA *108*, 15516–15521.
48. Hodgson, M., Steininger, G., and Razavi, Z. (2007). Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments. J. Acoust. Soc. Am. *121*, 2023–2033.
49. Plomp, R., and Mimpen, A.M. (1979). Speech-reception threshold for sentences as a function of age and noise level. J. Acoust. Soc. Am. *66*, 1333–1342.
50. Davis, A.C. (1989). The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. Int. J. Epidemiol. *18*, 911–917.
51. Moore, B.C.J., and Sek, A. (1996). Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. J. Acoust. Soc. Am. *100*, 2320–2331.
52. Bregman, A.S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound (The MIT Press).
53. Ferguson, M., and Henshaw, H. (2015). How does auditory training work? Joined-up thinking and listening. Semin. Hear. *36*, 237–249.
54. Mishra, J., de Villers-Sidani, E., Merzenich, M., and Gazzaley, A. (2014). Adaptive training diminishes distractibility in aging across species. Neuron *84*, 1091–1103.
55. Mishra, J., Anguera, J.A., and Gazzaley, A. (2016). Video games for neurocognitive optimization. Neuron *90*, 214–218.
56. Gervain, J., Vines, B.W., Chen, L.M., Seo, R.J., Hensch, T.K., Werker, J.F., and Young, A.H. (2013). Valproate reopens critical-period learning of absolute pitch. Front. Syst. Neurosci. *7*, 102.
57. Engineer, C.T., Hays, S.A., and Kilgard, M.P. (2017). Vagus nerve stimulation as a potential adjuvant to behavioral therapy for autism and other neurodevelopmental disorders. J. Neurodev. Disord. *9*, 20.
58. Song, J.H., Skoe, E., Banai, K., and Kraus, N. (2012). Training to improve hearing speech in noise: biological mechanisms. Cereb. Cortex *22*, 1180–1190.
59. Henderson Sabes, J., and Sweetow, R.W. (2007). Variables predicting outcomes on listening and communication enhancement (LACE) training. Int. J. Audiol. *46*, 374–383.
60. Papakonstantinou, A., Strelcyk, O., and Dau, T. (2011). Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise. Hear. Res. *280*, 30–37.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
CL Auditory Training Software	This paper	N/A
Auditory Memory Training Software	This paper	N/A
Mobile Auditory Testing Platform	This paper	N/A
MATLAB	https://www.mathworks.com/	N/A
R	https://www.r-project.org/	N/A
Unity Game Engine	https://unity3d.com/	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Jonathon Whitton (jonathon_whitton@harvard.meei.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subject Recruitment

All procedures were approved by the Human Studies Committee at the Massachusetts Eye and Ear Infirmary and the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology. Informed consent was obtained from each participant. For this study, we recruited 50 to 89 year old adults (16 male, 32 female) with mild to severe sensorineural hearing loss who used hearing aids full-time in both ears. Most subjects were referred to this study by their clinician.

METHOD DETAILS

Subject enrollment and training procedures

Forty-eight individuals were consented into this study and completed a screening visit to assess intelligence (Wechsler's Abbreviated Scale of Intelligence II), cognition (Montreal Cognitive Assessment), depression (Geriatric Depression Scale), medical history (103 item questionnaire), pure-tone detection thresholds (audiologist administered in sound treated booth), and hearing aid performance. To be included in the study candidates were required to be between 50 and 89 years of age, with mild to moderately-severe binaural sensorineural hearing loss (250-4000 Hz). Candidates were also required to be full-time bilateral hearing aid users for at least six months, be native English speakers, and present with a Full-scale IQ > 80, a Geriatric Depression Scale < 6, and a Montreal Cognitive Assessment > 25. Candidates were excluded if they reported a history of head injury, heavy alcohol consumption, neurological disorder, or current use of psychotropic medications. Thirty-six adults met inclusion/exclusion criteria based on screening measurements and were invited to return for baseline assessment and stratified randomization into the study (Figure S1). Hearing aid fit verification was performed by placing a small microphone in the ear canal, measuring the output of the hearing aid to various signal levels, and comparing the acoustic output of the hearing aid to gain prescriptions (NAL-NL2) based on the individual's hearing loss from 250-4000 Hz (Audioscan, Axiom system). Sixty-nine percent of the subjects' hearing aids were fit within 5 dB of the prescriptions and the remainder was between 5 and 10 dB of prescribed gain. Verification equipment was unavailable to make measurements for four subjects. Thirty-two individuals returned to the clinics to undergo baseline assessment. Twenty-four participants completed baseline assessments, training, and post-assessment. Twenty-one of these subjects also completed the final two-month follow-up testing session. Thus, eight subjects began the study but did not complete training or post-assessment. Of these, two dropped out before completing the initial baseline assessment. Four individuals dropped out after completing the baseline assessment and a few days of training. Two subjects were dismissed by the study leader due to lack of compliance over the first month of the study.

Of the 24 individuals who completed the pre- and post-testing sessions, 11 participants were randomly assigned to play the auditory memory game (45% female, mean age = 70 years ± 11) and 13 participants were randomly assigned to train on the CL audio tracking game (77% female, mean age = 70 years ± 7). More than half of the participants in each training group reported formal musical training (64% of memory training group and 54% of CL training group), though none identified themselves as past or current musicians. Most participants were college graduates (73% of memory training group and 77% of CL training group). Two participants in the CL training group reported that they were bilingual.

Subjects performed all behavioral testing and game training on individually assigned Microsoft Surface Pro 2 tablets. The interactive software used for both games and the behavioral testing was developed as a Windows Store App using the Unity game engine and side-loaded onto the tablets. Audio stimuli were presented through a Dell AX 210 speaker that was connected to the tablet and placed at an approximate distance of 1 m and azimuthal position of 0° relative to the subject. Subjects performed a calibration at each home testing/training session to ensure reliable positioning of the speaker throughout this four-month study. Briefly, the calibration program launched the native front-facing camera on the tablet and provided the subjects with visual guides to adjust the position of the speaker relative to the tablet device. Once the speaker was properly placed, the subject touched the screen to transmit an image of the test/training setup to the secure servers at MEEI for offline review. The native microphone on the tablet was also used by the custom application to make ambient noise level measurements in the home environment. If noise levels exceeded 60 dB A, the participant was locked out of the software, provided with a warning about excessive noise levels in the test environment, and prompted to find a quieter location for testing. Average ambient noise levels measured by the tablet were 43 ± 4 dB A. Subjects used their experimental tablets to establish a wireless internet connection from their home environments. Data collected during testing and gameplay were automatically encrypted and uploaded using a secure file transfer protocol. Subjects did not receive compensation for their participation in this study, though parking fees were reimbursed and we did enter participants in a drawing to win one of five tablets at the conclusion of the study.

Subjects were randomly assigned to play the auditory memory game or the CL audiomotor game for two months. Both tasks were embedded in a puzzle game. Subjects earned puzzle pieces by successfully executing their assigned tasks and were able to use them to reconstruct paintings by well-known artists or their own photos. The skins and graphics for training game environment were identical. Group randomization was stratified based on the subjects' baseline performance on the Quick Speech-in-Noise test. This stratification was performed because baseline performance on speech recognition tests is reported to have prognostic value in speech training studies [58, 59], though we found no evidence of this in our generalization study [30]. Group randomization was automated by an algorithm that cumulatively tracked baseline sentence-in-noise scores for randomized subjects. Therefore, members of the research team were not involved in randomization and did not know which game the subjects would play. Subjects were only exposed to the game that they were randomly assigned to play, and all instructions for gameplay and testing were provided through video tutorials as well as assisted play in the application. In this way, all aspects of the daily testing and training activities in which the subjects engaged were addressed by the user interface of the software application. However, four subjects (two from each game group), had trouble using the training interfaces from home. These subjects returned to the clinic and were provided with a 30-min coaching session by a study staff member. Importantly, these staff members were not involved in data scoring or analysis, allowing us to maintain double-blinding during the study. Individuals were asked to train on their respective game for 3.5 hr each week. To count toward their weekly goal, each session was required to last at least 30 min, but no longer than 1 hr. In this way, the participant could complete 30 min sessions every day of the week or, at the opposite end of the spectrum, they could complete three one hour and 1 half hour session each week. Both training applications provided visualizations of each participant's progress toward his/her weekly goal, allowing participants to plan when/how to complete their training time. Participants were free to train during any time of day that was convenient for them; at baseline, all participants were asked to train during a time of day that they could find a quiet place and focus for at least 30 min. Presentation levels for game sounds were tailored to the aided hearing sensitivity of each participant. Specifically, baseline audiograms and loudness discomfort levels were measured for carrier frequencies between 125 and 8000 Hz with hearing aids on. These threshold and loudness discomfort measurements were interpolated, and game stimuli were presented at 20-40 dB SL with the restriction that stimulus values did not reach the measured level of discomfort. This ensured that sounds in the game were audible and comfortable in a frequency-specific manner, tailored for each participant in the study.

Auditory memory game

We developed an auditory memory game as a control intervention for this study. The choice to develop a speech-based auditory memory game was based on our expectation that it would have good face-validity as an intervention for speech recognition abilities, it would promote auditory memory learning [39–41]. During the task, subjects heard one or more strings which took the form of "Ready *name* go to *color number* now." The *name*, *color*, and *number* for each string were randomly selected from eight, four, and eight possibilities respectively. After the subjects were presented with the auditory string, a number of labeled virtual elements slowly emerged on the screen, each with a non-overlapping 0.5 to 1.5 s delay. The subjects' task was to identify virtual elements on the screen that corresponded to the *name, color, and number* that they had previously heard and then to connect these elements to create a composite object. Game difficulty adaptively changed by incrementing the number of distractor elements, the speed of elements, the number of phrases that were spoken, or the number of phrases that required responses. We also administered "yardstick" conditions periodically throughout the study to track learning under the same perceptual and cognitive demands (two name-color-number strings presented). We used publically available recordings of the Coordinate Response Measure corpus generated by eight different speakers to create all of the task-related auditory stimuli for this game.

Each time that a player successfully matched a *name-color-number* string, they were then required to position it in a target location in order to generate a new puzzle piece. After they had generated a sufficient number of pieces, they were taken to a new sub-game screen where they were asked to use the pieces that they had earned to construct a puzzle (Figure S3). As with any jigsaw puzzle, the subjects were provided with patterned and geometrical cues for the correct positioning of each puzzle piece. To provide feedback for the subjects' positioning of the virtual pieces, the tiles would "snap into place" when they were placed in the correct location. While subjects performed this spatial reasoning task, they were simultaneously presented with stimuli from the CL audiomotor game

(see below). Specifically, they heard the audio generated by a random search in an auditory gradient. Because the sound was not tied to their motor activities, the participants generally perceived it as a distractor stimulus, and it served as a control condition for passive exposure to the same stimuli used in the audiomotor learning game.

CL audiomotor game

The audiomotor training task used in this study was initially inspired by sensory-guided foraging behaviors in rodents and refined based on our own experiments involving CL audiomotor learning [30]. The basic CL audiomotor method involves the establishment of an acoustic gradient that is mapped to some physical or virtual space. Subjects explore the space and their searches reveal the manner in which their motor behaviors parametrically alter the stimulus attributes. This information can then be iteratively used by the forager to identify hidden spatial targets more efficiently [30].

In the auditory tracing task used in this study, subjects were aware that the outline of a polygon was hidden somewhere on the screen. They were required to use either a stylus or their finger to identify the location of the polygon and trace the outline of its shape. An auditory gradient was established relative to the individual lines comprising the edges of the shape, and as a subject moved his/her stylus though the gradient, either the level, frequency, or modulation rate of the sound was changed logarithmically with the subject's distance from the shape outline. Subjects adaptively learned to use these real time cues to reveal the shape with less error over the course of training (Figure 1E) Game difficulty increased over the course of the study in two ways. The signal to noise ratio adaptively changed based on the subjects' performance, such that its maximum value decreased after every puzzle completion. Additionally, the complexity of shapes (defined by number of vertices) was increased following each puzzle completion. We also administered "yardstick" conditions periodically throughout the study to track learning under the same perceptual demands (−18 dB SNR atop the shape outline). The dependent measurement that we used to define learning on the yardstick conditions was audio tracing error, which is defined as the perpendicular Euclidean distance between a players current position and the nearest line segment, Tracing error = $\sqrt{(traceX - lineX)^2, (traceY - lineY)^2}$.

Based on the accuracy of the subject's tracing, they were awarded time to complete two sub-games. The first sub-game was a gradient-based search task, similar to that used in our previous training studies in rats, mice and young adult human subjects with normal hearing [30–32] (Figure S4A). Subjects were required to move the puzzle piece to a target location on the display. The target location was invisible, but a circular audio gradient was established that logarithmically varied audio stimulus attributes with the instantaneous Euclidean distance from the target. Once the subject thought that they had found the correct location, they would release the virtual puzzle piece. If the subject was correct the piece would remain in place, but if they were incorrect (outside the rewarded area), the piece would fall to the bottom of the screen and a new target area and gradient would be randomly generated. If sufficient time remained on the countdown timer, the subject began a second sub-game that challenged them to rotate the virtual puzzle piece around a central axis to achieve the correct orientation (like a combination lock, Figure S4G). There were no visual cues concerning the correct orientation of the puzzle piece, but the audio stimulus would rapidly change its value from a reference sound to a target sound when the piece was rotated into the correct orientation. If the subject rotated beyond the correct point or released the piece prior to the correct point, a new target orientation was selected and the user was permitted to try again until the time expired.

The three sub-games were packaged as a group into sound "worlds" defined by the three sound features that subjects were asked to discriminate: pitch, level, and amplitude modulation. The stimuli used for each world consisted of amplitude modulated pure tones, spectrotemporally modulated ripple noise, and tone clouds. The levels of all stimuli varied from 20-40 dB sensation level (dB SL) and were limited by the measured loudness discomfort levels of each subject. Minimum sensation level and maximum tolerable level were defined for each subject across a range of pure tone frequencies prior to the start of training. The carrier frequencies of tones varied from 125-8000 Hz. The modulation rates of tones varied from 2-32 Hz. Spectrotemporally modulated ripple noise was synthesized from sinusoidal components with frequencies spaced from 354-5656 Hz in 0.05-octave steps. Ripple density varied from 0.5 to 3 cycles per octave, and modulation velocity varied from 4-12 Hz. For the tone clouds, 50 ms tone pips were randomly selected from a uniform distribution that varied in bandwidth from 0.25 to 1.5 octaves. The level of each tone pip in the cloud was roved by ± 6 dB. The dynamic range provided across the full extent of the spatialized sound gradient was 40 dB for worlds that focused on level cues, 4 octaves for worlds that focused on frequency cues, and 0.125-12 Hz for worlds that focused on ripple velocity cues. All game signals were presented while 1-6 talker babble played in the background. Background speech materials were generated by concatenating a subset of IEEE sentences (sentences 361-720) presented by 20 different talkers from the Pacific Northwest/Northern Cities corpus. The composition of the babble speakers was randomly selected and presented for the entirety of a given game board. The loudness of the sentences was balanced prior to concatenation. The subset of IEEE sentences that we used as distractors in this study were selected such that they did not overlap with the IEEE sentences used in the Quick Speech-in-Noise Test (sentences 1-360).

Psychophysical tasks to assess transfer of learning

All behavioral testing was self-directed. Subjects interacted with a custom software interface to perform alternative forced choice, reaction time, and open response tasks. Subjects wore their hearing aids and were asked to use their typical settings for all testing and training performed in the study. Each behavioral task began with instructions and practice trials. In the practice trials, the perceptual demands were kept at an "easy" level and subjects were given feedback concerning the accuracy of their responses (with the exception of speech recognition tasks). Subjects were required to achieve a minimum performance level to assure that basic procedural aspects of the task were learned before the testing blocks began. In previous experiments, we found that home testing using this software interface provided results that were statistically equivalent to manual testing in sound treated rooms [43].

Speech recognition in noise

For all speech recognition in noise tasks, subjects were asked to repeat a target talker who produced either a word or a sentence. After the speaker finished, the subjects touched a virtual button on the tablet screen to activate the tablet's native microphone and record their verbal responses. These responses were saved as encrypted raw binary files, transmitted wirelessly to secure servers, converted to .wav files, and scored offline by a blinded experimenter. Word recognition in noise testing was conducted using the Words In Noise test (WIN). The WIN is a clinical test that consists of monosyllabic words from the Northwestern University 6 corpus spoken by a female talker while 6-talker babble was played continuously in the background. 35 monosyllabic words were presented at SNRs that varied from 24 to 0 dB SNR in 4 dB steps. We administered a unique randomization of WIN lists 1 and 2 at each time point in the study.

Sentence recognition in noise was assessed using two clinical tests, the Quick Speech-in-Noise test (QuickSiN) and the BKB Speech-in-Noise Test (BKBSiN). For both the QuickSiN and the BKBSiN, target sentences were presented while 4-talker babble played continuously in the background. These speech corpora differ in the characteristics of the target speaker's voice, the number of keywords in a given sentence, and linguistic load. The difference we highlight here is that QuickSiN employed sentences with lower linguistic context (IEEE sentences) while the BKBSiN test used sentences that contained higher linguistic context (Bamford-Kowal-Bench sentences). The signal to noise ratio of the QuickSiN varied from 25 to 0 dB SNR in 5 dB steps, while the signal to noise ratio of the BKBSiN varied from 21 to −6 dB SNR in 3 dB steps. Four unique sentence lists (QuickSiN) or two unique list pairs (BKBSiN) were administered at each testing time point in the study. Additionally, two list pairs of the BKBSiN were measured while the subject was not wearing hearing aids during the pretest visit to establish the amount of benefit provided by the subjects' hearing aids (Figures 4A and 4B). The lists that were used for aided and unaided testing were randomly selected, as was their presentation order.

Memory Assessment

Subjects were tested with the Letter Number Sequencing test (Wechsler's Adult Intelligence Scale III). All audio stimuli were pre-recorded and shared with us by Dr. Adam Gazzaley's laboratory at the University of California, San Francisco [21, 54]. Audio stimuli consisted of increasingly long strings of letters and numbers spoken by a male talker. An experimenter at our research facility administered the test in a clinical sound booth. The experimenter initiated each trial with a virtual button press on the tablet. After the stimulus presentation, the native microphone of the tablet was used to record the responses of the subjects. The subjects were asked to verbally respond by repeating all elements of the string with the numbers first in ascending order, followed by the letters in alphabetical order. The experimenter scored the participants' responses online in order to determine when testing was to terminate (following 3 incorrect responses at the same memory load level). However, the actual scoring of the data that are presented in this manuscript was performed by a blind experimenter after the .wav files of the participants' responses were uploaded to our servers. It should be noted that due to the subjects' deficient speech processing and the auditory-only presentation mode of the stimuli, subjects often made phonemic confusions, even under conditions of low memory load. For this reason, we scored their responses in two ways, strict and loose. For strict scoring, any phonemic mistake was counted as incorrect. For loose scoring, confusions that involved up to two of the distinctive features of phonemic categories were tolerated as hearing errors (e.g., place and voice onset time errors). Training effects in the study were nearly identical with either scoring method. Data are reported here with the loose scoring method.

Competing Digits

Subjects were initially familiarized with a male speaker (fundamental frequency = 115 Hz) as he spoke 120 digits in relative quiet. On each trial, the male speaker produced a string of four randomly selected digits (digits 1-9, excluding the bisyllabic '7') with 0.68 s between the onset of each digit stimulus. The subjects used a virtual keypad on the tablet to enter with the digits spoken by the target speaker. After familiarization, two additional talkers were introduced (male, fundamental frequency = 90 Hz; female, fundamental frequency = 175 Hz) as distractors. These distractor speakers also produced randomly selected digits with target-matched onset times. The only contingency was that two speakers could not produce the same digit at once, otherwise the digit produced by each speaker was selected at random. The target speaker was presented at 65 dB SPL. Four hundred and twenty eight digits were presented at 0 dB SNR (target and distractors at the same level), and ninety two digits were presented at 3 dB SNR (the target was 3 dB higher in level than the distractors). We observed that 32% of the subjects performed at chance level in the more challenging 0 dB SNR condition during baseline testing. By contrast, only 8% of the sample performed at chance levels in the 3 dB SNR condition. To avoid floor effects in our analysis, we focused on the 3 dB SNR condition. We analyzed performance on the digits task in two ways. First, we asked how often the subjects correctly identified the first digit in each stream. We viewed performance on this condition as analogous to monosyllabic word recognition in noise task. Next, we asked how often the subjects correctly identified all four digits in a stream. We viewed performance under this condition as analogous to a sentence recognition in noise task.

Audio/Visual Stroop

The Stroop effect provides a well-established measure of inhibitory control. For all versions of the Stroop tasks, subjects are asked to attend to a stimulus and then report the identity of the attended attribute while ignoring irrelevant stimulus attributes. In some cases, the "distractor" stimulus attributes are congruent with the target stimulus attribute and in other cases they are incongruent. The congruency of target and distractor stimulus attributes has a marked effect on reaction times (RT) with responses to congruent conditions occurring ~250 ms sooner than responses to incongruent trials. A neutral condition is also presented wherein there is no congruency relationship between the target and distractor stimulus features. The neutral condition provides a control measurement for processing speed and can be used to compute normalized Stroop interference $(\mathit{Incongruent\ RT}(s) - \mathit{Congruent\ RT}(s)/\mathit{Neutral\ RT}(s))$.

We measured performance on a visual and audio Stroop task in this study. In the visual Stroop task, subjects were visually presented with the text, "Red', "Blue," and "Legal" in a random vertical location on the screen (letter height = 3.5 cm, white background). The color of the word was either red or blue. This created three conditions, color-letter congruency, color- letter incongruency, and a neutral condition (the word "Legal"). Likewise, the audio Stroop employed three words ("High," "Low," and "Day") that were either spoken with a low fundamental frequency (180 Hz) or a high fundamental frequency (280 Hz). The three words were spoken by the same female talker, and the TANDEM-STRAIGHT vocoder was used to synthesize these three vocalizations and shift the fundamental frequency up and down to create high and low pitch versions of each word. Subjects began each trial by placing their thumbs in two circles that were positioned on each side of the capacitive touch screen, midway along the vertical axis. After a 0.5-2 s delay, an audio or visual stimulus was presented and two virtual response buttons appeared just above and below each thumb fixation circle. The participants were required to select one of the two responses as quickly and accurately as possible. Their reaction times were recorded as the latency of the first of the two thumb responses.

Before each trial began, either a visual or an audio masker was presented to cue the trial and wash out stimulus recency effects. The visual masker was a grid of 39 individually colored squares (grid dimensions 16.9 × 4.2 cm) positioned 3.6 cm from the top of the tablet screen. The color of each element in the grid was randomly selected to be red, blue, green, or yellow with a refresh rate of 4 Hz. The audio masker consisted of 15 tones that were each 50 ms in duration and presented with an interstimulus interval of 0.18 s at a level of 60 dB SPL. The carrier frequency of each tone was randomly selected from an interval of values that ranged from 500 to 8000 Hz. The duration of the video and audio maskers were 1 s each. To compute average reaction times for the congruent, incongruent, and neutral conditions, each word-color and word-pitch combination was repeated 30 times over the course of 3 training blocks. Testing was complete once each individual had accrued at least 40 correct responses for each of the three congruency conditions (i.e., incongruent, congruent, and neutral). The reaction time of correct responses was analyzed.

Frequency Modulation Detection

The subjects were initially exposed to the perceptual experience of frequency modulation (FM) through an interactive slider that they manipulated to increase and decrease the excursion depth of a frequency modulated tone. High excursions were labeled as 'squiggly' to allow the subjects to associate the sound with a label that could be used when completing the 2-interval 2-alternative forced choice FM detection task. After initial familiarization, two tones (carrier frequency = 1000 Hz, duration = 1 s, level = 55 dB SL) were presented to subjects with an interstimulus interval of 0.5 s. Frequency modulation was applied at a rate of 2 Hz to one of the two tones (random order). A quasi-sinusoidal amplitude modulation (6 dB depth) was applied to both tones to reduce cochlear excitation pattern cues [60]. The subject was asked to indicate whether the first or second tone was frequency modulated ('squiggly'). The two-down-one -up procedure was used to modulate the frequency excursion magnitude in order to converge on the 70.7% correct point. The frequency excursion of the FM tone was initially set to 75 Hz and changed by a factor of 1.5 for the first 5 reversals, decreasing to a factor of 1.2 for the last 7 reversals. The geometric mean of the last 6 reversals was used to compute the run value. A minimum of 3 runs were collected. The coefficient of variation across runs was computed online. If the coefficient of variation was > 0.2, additional runs were collected until this criterion was met or six runs had been collected, whichever came first. The median threshold across all runs collected was used to define the participant's FM detection threshold.

Analysis of speech reception data

We primarily analyzed the speech recognition data by computing correct scores over challenging SNRs for each test. Across both groups, performance changed as a function of noise level over a similar range of SNRs. We used SNRs from the steeply sloping portion of the psychometric function to assess performance (WIN, 12-16 dB SNR; QuickSiN, 5-10 dB SNR; BKBSiN, 0-6 dB SNR). Scores were expressed as rationalized arcsine units (RAU) since percentage or proportional scores generally violate several assumptions of parametric statistical tests for values near zero and one. We also performed clinical scoring for each speech test by computing a non-adaptive threshold using the Spearman-Kärber equation (Figure 4D). To summarize our findings for the words and sentence tests, we computed a change index because zero values in a few of the full digit sequence tests made the ratiometric analyses used for other speech tests impossible, $(Post\ score(\%) - Pre\ score(\%)/Post\ score(\%) + Pre\ score(\%))$. We used the change index to compare the patterns of results between the open-set speech tasks and the digit streaming task. We used Hedges' *g* to compute treatment effect sizes for the speech recognition and digits streaming tasks at the post and two-month follow-up assessment periods. We computed 95% confidence intervals around the effect size using a bootstrapping approach.

QUANTIFICATION AND STATISTICAL ANALYSIS

Normality of data distributions was assessed using the Shapiro-Wilk test and q-q plots. In the cases of normal distributions, parametric tests were employed to test for statistical significance. Statistical significance of group intervention effects was assessed by performing repeated-measures, mixed effects, and multivariate ANOVA. Analysis of covariance (ANCOVA) using the subjects' pretest scores as the covariate on the outcome measures provides the same pattern of significant results reported in the manuscript (Table S1). Other between group comparisons were tested using two-sample t tests. All statistical analyses were two-tailed. Group comparisons between non-normally distributed data were made using the Wilcoxon rank sum test. Correlations were quantified using Pearson's linear correlation coefficient and corrected for multiple comparisons when appropriate (Holm-Bonferroni). Prior to initiating the study, a power analysis using the Hotelling-Lawley Trace statistic suggested that the experiment would be adequately powered ($\beta = 0.2$) to test the null hypothesis ($\alpha = 0.05$) with a sample size of 28 participants; we enrolled 32 participants but early dropouts reduced the sample size to 24. Nevertheless, the study was still adequately powered to test the null hypothesis because

the measured placebo effects in our control group were lower than our estimate and the pre-and post-test outcome measure correlations were higher than we had anticipated.

Initial data processing, visualizations, and summary statistical descriptions were performed using custom scripts written in MATLAB Repeated-measures ANOVAs and t tests were also performed using MATLAB scripts. Repeated-measures ANCOVA and mixed effects ANOVA were performed in R. Outcomes of most statistical tests are reported in the Results section of the main text. Additional tests and summary statistics for the primary outcome measures have also been reported in Tables S1 and S2. Reported samples sizes (N) throughout the manuscript refer to number of participants.

DATA AND SOFTWARE AVAILABILITY

Deidentified datasets associated with outcome measurements and custom MATLAB analysis scripts will be made available upon request to the Lead Contact, Jonathon Whitton (jonathon_whitton@harvard.meei.edu).

ADDITIONAL RESOURCES

This study was registered on ClinicalTrails.gov (Identifier NCT02147847) under the title ‘‘Computer-Based Auditory Rehabilitation’’ on May, 15 2014.

Current Biology, Volume *27*

Supplemental Information

Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise

Jonathon P. Whitton, Kenneth E. Hancock, Jeffrey M. Shannon, and Daniel B. Polley

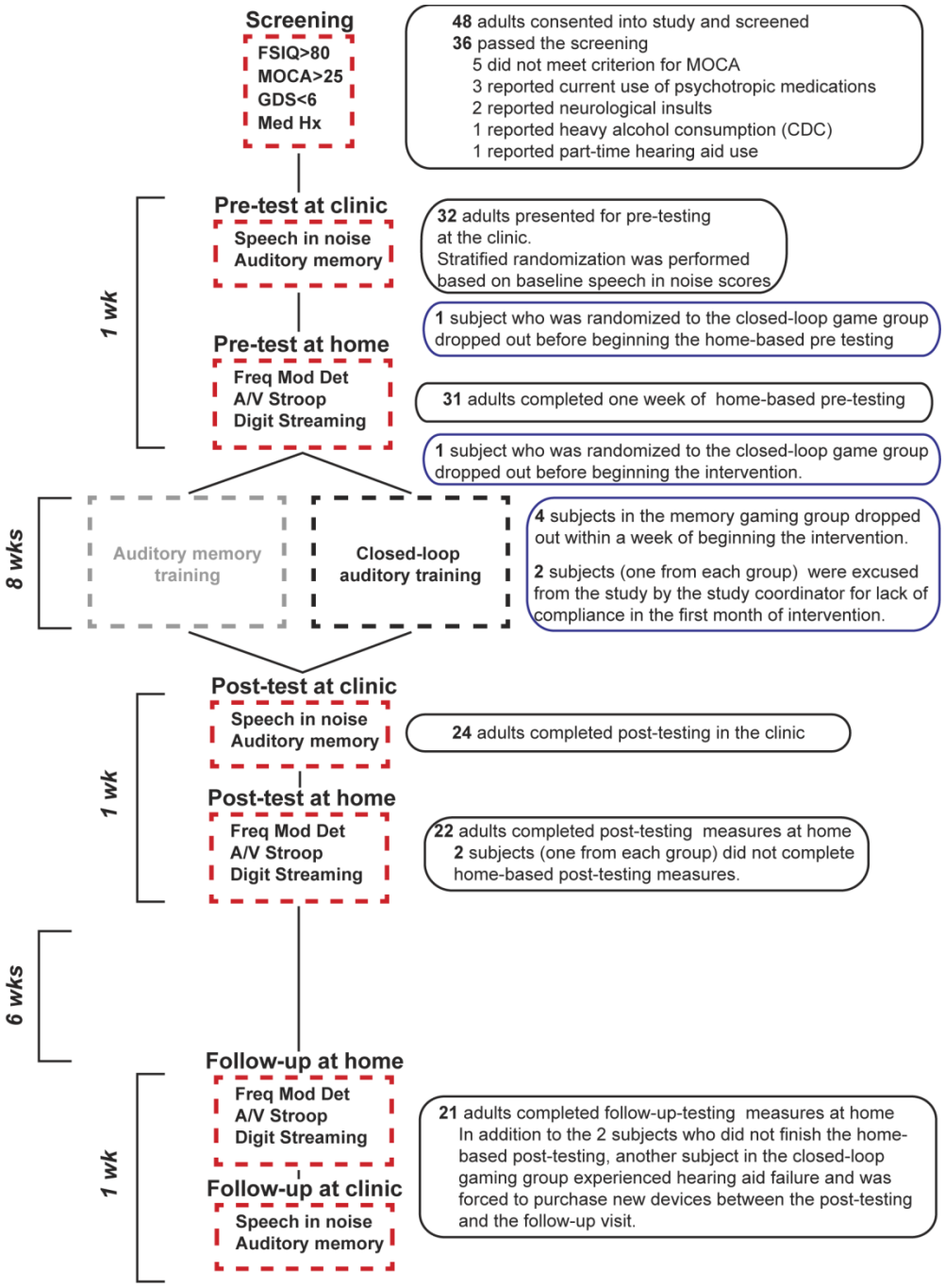

Figure S1 Study flowchart, related to Figure 1. (**A**) Of the thirty-two participants who began the study, twenty-four completed the pretest, intervention, and posttest. Twenty-one participants completed the 2 month follow-up assessment.

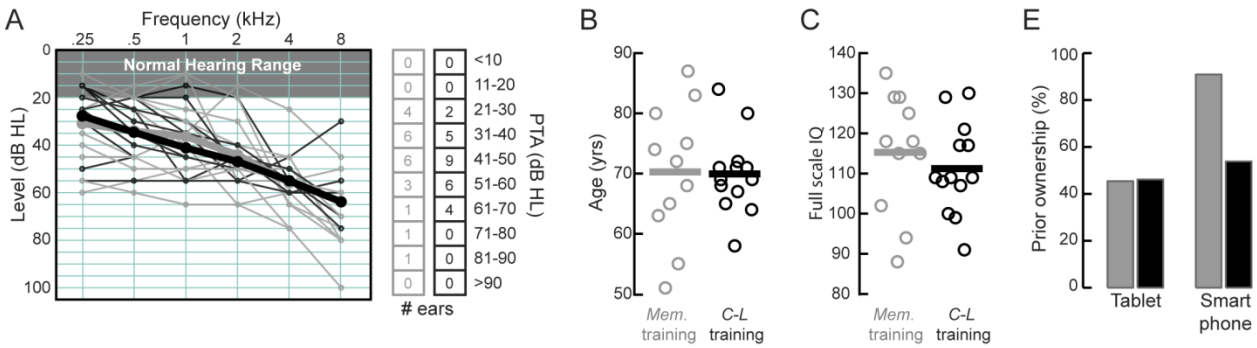

Figure S2 Study demographics, related to Figure 1. (**A**) Air conduction pure tone detection thresholds were collected by an audiologist in a sound treated booth. Subjects generally presented with mild sloping to moderately-severe sensorineural hearing loss (working memory game = gray, closed-loop game = black). Distribution of pure tone averages (PTA .5 – 2 kHz) in the sample plotted according to AAO-HNS recommendations [S1] (*right*). (**B**) Participant age, (**C**) full scale IQ and (D), prior technology ownership were balanced across training groups.

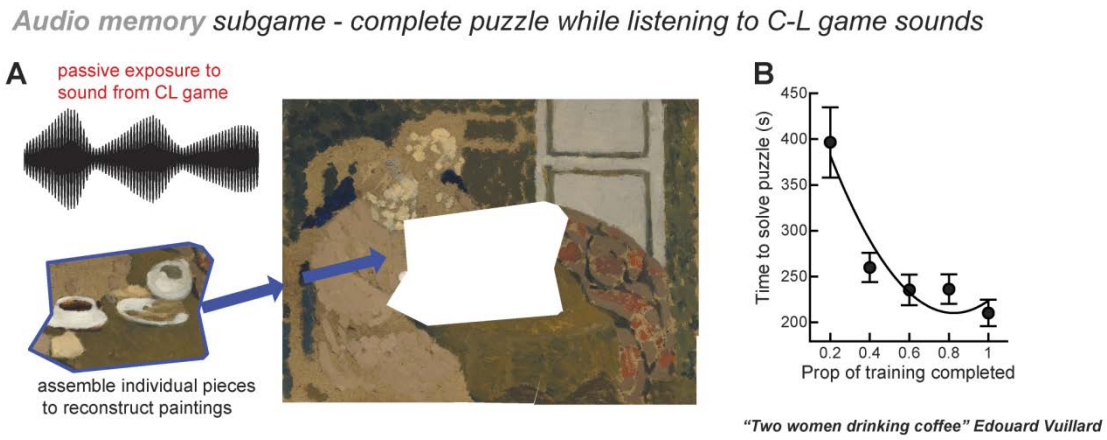

Figure S3 Sub-game design for the auditory working memory training task, related to Figure 1. (**A**) The auditory memory game involved a visuospatial puzzler sub-game. Initially, the puzzle pieces were jumbled at the bottom of the screen. The participant touched a puzzle pieced with their finger, guided it to the correct position on the puzzle board, and then rotated it into the correct orientation. While the subject performed this task, dynamic auditory stimuli from the closed-loop training game were played in the background (black time waveforms, top left). So, though the motor behaviors and the auditory stimuli in this sub-game were the same as the CL training (**Figure S4**), the motor behavior not linked in a CL to the auditory stimuli they were hearing. Instead, puzzles were completed using visuospatial cues. (**B**) Subject performance improved substantially on this puzzle sub-game; the time required to complete a puzzle reduced by nearly 50% over the course of training (F = 15.2, P = 1.2×10^{-7}, Repeated Measures ANOVA).

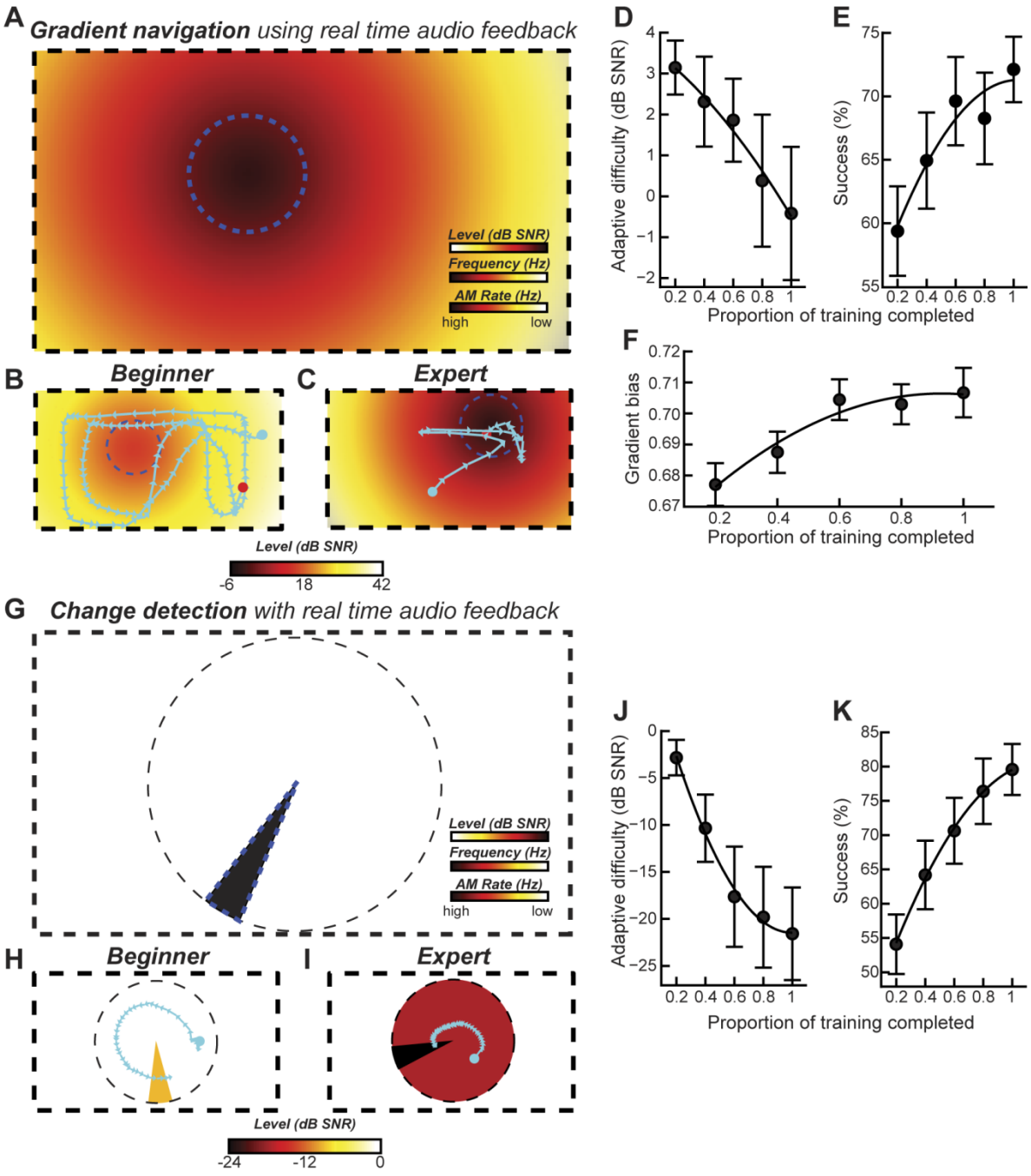

Figure S4 Sub-game design for the closed-loop audiomotor task, related to Figure 1. There were two closed-loop audio sub-games. (**A**) The first game involved navigating audio gradients that varied logarithmically with distance from a circular target in order to place a puzzle piece in its correct location. (**B-C**) While beginners attempted to solve the task through exhaustive searches (*B*), expert players used the most information rich vectors, biasing their searches along the highest sloping portions of the gradient (*C*). Cyan arrows depict the subject's paths, with each arrow representing movement over a

0.5 s time bin. (**D**) Difficulty in the game adaptively increased via reduction of the SNR (D, $F = 3.72$, $P = 0.01$, RMANOVA). (**E**) By examining performance on “yardstick” trials (where SNR was fixed at -18 dB), we observed significant success rate improvements with training (E, $F = 6.8$, $P = 2.0 \times 10^{-2}$, RMANOVA). (**F**) Using the “yardstick” trials, we also examined the degree to which subjects aligned their search vectors with the most informative region of the gradient ($Gradient\ bias = \frac{v \cdot w}{(|v||w|)}$, where v is the player’s actual traveled angle and w is the angle between the player and the target). We observed significant increases in gradient bias with training ($F = 8.4$, $P = 3.0 \times 10^{-5}$, RMANOVA). Gradient bias values range from 0.64 (chance) to 1 (movements perfectly aligned with the highest sloping portion of the gradient). (**G**) The second sub-game involved rotating the puzzle piece around a central axis to identify the correct orientation. As the subject rotated the puzzle piece, sound level, pitch, or rate stayed constant until the target angle was reached, at which point the target feature was modulated with a step function. The subjects were required to stop rotation and release the puzzle piece immediately upon detecting this change. (**H-J**) Rotating beyond this point, as visualized in a trial generated by a beginner (*H*) resulted in failure and re-randomization of the target orientation. Likewise, releasing the piece before arriving at the target location also resulted in failure. Expert users (*I*) learned to perform this task accurately at progressively worse SNRs (*J,* $F = 18.2$, $P = 3.0 \times 10^{-9}$, RMANOVA). (**K**) By examining performance on “yardstick” conditions (where SNR was fixed at -18 dB), we observed significant success rate improvements with training (K, $F = 16.57$, $P = 1.0 \times 10^{-8}$, RMANOVA).

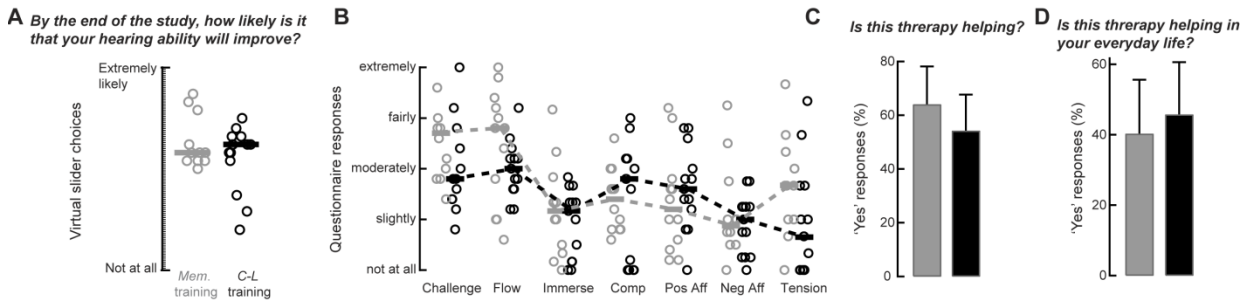

Figure S5 Expectations and game play experience were matched between the auditory memory and closed-loop audiomotor training tasks, related to Figure 1. (**A**) After subjects played their training game for a week they were asked to use a virtual slider to rank their expectancy that their hearing would improve as a function of playing their assigned game. Expectations were well matched across training groups (gray = memory game, black = closed-loop game, $z = 0.18$, $P = 0.86$, Wilcoxon rank-sum). (**B**) At the same time point, we also measured the participants' impressions of their game experience using the Game Experience Questionnaire [S2, S3]. Responses to questions on the Game Experience Questionnaire are divided into seven experience categories. Both games were rated as moderately to fairly challenging and only slightly immersive. Flow, which involves questions concerning "losing track of time" and "being fully occupied with the game," was also ranked as moderate to fair. (Comp = competence, Pos Aff = positive affect, Neg Aff = negative affect). There were no significant differences between the ratings that the memory and closed-loop games received across categories. The largest non-significant difference between the two groups was found for challenge, with the WM task being rated as more challenging than the CL task ($z = 1.75$, $P = 0.08$ uncorrected, $P = 0.56$ Holm-Bonferroni corrected for multiple comparisons, Wilcoxon rank-sum). (**C**) After the subjects had trained for 1 month, they were asked whether they felt that the therapy was helping. About half of each group responded affirmatively to that question. (**D**) After two months of training we asked participants if the therapy was helping in their everyday lives. Around 40% responded affirmatively to this question. There were no significant response differences between groups for either of these questions ($P \geq 0.7$, Fisher's exact test).

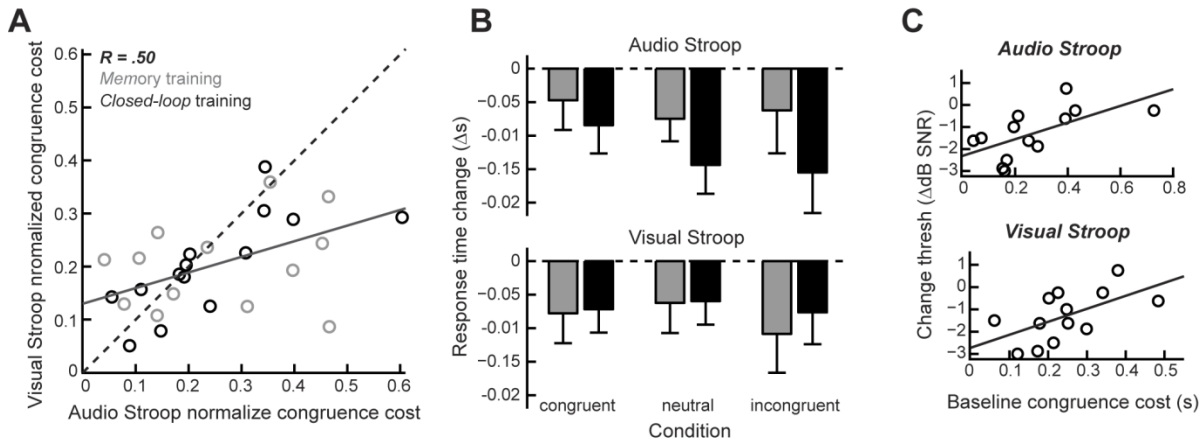

Figure S6 Estimating cognitive interference with audio and visual Stroop tests, related to Figure 5. (**A**) The normalized congruence cost for the visual and auditory versions of the Stroop task were significantly correlated (R = 0.50, P = 8 x10^{-3}, Pearson's correlation coefficient, gray = WM training, black = CL training). (**B**) Reaction times decreased for all congruency conditions ($P \leq 0.05$ for all conditions, time effect, RMANOVA). (**C**) Baseline performance on both the audio and visual versions of the Stroop task were significant predictors of learning transfer to the sentence recognition in noise tasks following closed-loop training (C, Audio: R = 0.62, P = 0.048; Visual: R = 0.58, P = 0.04, Pearson's correlation coefficient, Holm-Bonferroni corrected for multiple comparisons).

Analysis of Covariance: Pre vs Post scores

Outcome measure	*F* value	*P* value
Words in noise	0.3	0.60
Low context sentences in noise	***5.2***	***0.03***
High context sentences in noise	***5.1***	***0.04***
Frequency modulation detection	0.0	0.96
Letter number sequencing	0.4	0.65
Audio stroop	0.7	0.40
Visual stroop	0.2	0.67

Table S1 Analysis of Covariance, related to Figure 2. An alternate analysis of changes in outcome measures following training was executed by performing ANCOVA using the baseline score as a covariate. The pattern of statistical significance is identical to that obtained using interaction terms of the repeated measures ANOVA reported in the main text.

Supplemental References

S1. Gurgel, R.K., Jackler, R.K., Dobie, R. a., and Popelka, G.R. (2012). A new standardized format for reporting hearing outcome in clinical trials. Otolaryngol. -- Head Neck Surg. *147*, 803–807.

S2. Norman, K.L. GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers. Interact. Comput. *25*, 278–283.

S3. IJsselsteijn, W., van den Hoogen, W., Klimmt, C., de Kort, Y., Lindley, C., Mathiak, K., Poels, K., Ravaja, N., Turpeinen, M., and Vorderer, P. (2008). Measuring the experience of digital game enjoyment. In Measuring Behavior, A. Spink, M. Ballintijn, N. Bogers, F. Grieco, L. Loijense, L. Noldus, G. Smit, and P. Zimmerman, eds. (Maastricht), pp. 88–89.