

Training allows switching from limited-capacity manipulations to large-capacity perceptual processing

Tamar Malinovitch ¹, Philippe Albouy ², Robert J. Zatorre ³, Merav Ahissar ^{4,*}

¹Department of Cognitive and Brain Sciences, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 9190501, Israel,

²CERVO Brain Research Centre, Laval University, 2301 Av. D'Estimauville, Québec, G1V 0A6, Canada,

³Montreal Neurological Institute, McGill University, 3801, rue University Montreal, Québec, H3A 2B4, Canada,

⁴The Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem, The Edmond J. Safra Campus - Givat Ram, Jerusalem 9190401, Israel

*Corresponding author: The Edmond and Lily Safra Center for Brain Sciences, Hebrew University of Jerusalem, The Edmond J. Safra Campus - Givat Ram, Jerusalem 9190401, Israel. Email: msmerava@gmail.com

In contrast to perceptual tasks, which enable concurrent processing of many stimuli, working memory (WM) tasks have a very small capacity, limiting cognitive skills. Training on WM tasks often yields substantial improvement, suggesting that training might increase the general WM capacity. To understand the underlying processes, we trained a test group with a newly designed tone manipulation WM task and a control group with a challenging perceptual task of pitch pattern discrimination. Functional magnetic resonance imaging (fMRI) scans confirmed that pretraining, manipulation was associated with a dorsal fronto-parietal WM network, while pitch comparison was associated with activation of ventral auditory regions. Training induced improvement in each group, which was limited to the trained task. Analyzing the behavior of the group trained with tone manipulation revealed that participants learned to replace active manipulation with a perceptual verification of the position of a single salient tone in the sequence presented as a tentative reply. Posttraining fMRI scans revealed modifications in ventral activation of both groups. Successful WM-trained participants learned to utilize auditory regions for the trained task. These observations suggest that the huge task-specific enhancement of WM capacity stems from a task-specific switch to perceptual routines, implemented in perceptual regions.

Key words: perceptual learning; fMRI; frequency discrimination; pitch pattern; stimuli manipulation.

Introduction

The ability to keep accessible and manipulate recent information (working memory [WM]) is essential for most high-level cognitive processes, including reading, mathematical calculations, and problem-solving (Swanson 2004; Bayliss et al. 2005). Its capacity, measured as the number of items one can keep and manipulate in designated WM tasks, is very limited (~4 items, Cowan 2001) and is strongly correlated with measures of reasoning (Conway et al. 2003; Halford et al. 2007) and academic performance (Hitch et al. 2001). Functional imaging studies have found that brain regions along the dorsal stream (mainly in the posterior parietal cortex and the dorsal frontal cortex) are recruited when performing auditory (e.g. Klingberg 1998; Rodriguez-Jimenez et al. 2009; Zatorre et al. 2010; Foster et al. 2013; Albouy et al. 2017) or visual (D'Esposito et al. 1998; Champod and Petrides 2007, 2010) WM manipulation tasks.

The cognitive advantages associated with high WM capacity have led researchers and commercial companies to explore ways to elevate it by training. Indeed, although WM tasks require de novo manipulations, which challenge WM capacity, training often yields

impressive improvement (Holmes et al. 2009; Klingberg 2010). However, transfer is only found to very similar (“near”) tasks, and no transfer is found to “far” WM tasks (Shipstead et al. 2010, 2012; Melby-Lervåg and Hulme 2013; Melby-Lervåg et al. 2016; Meiran et al. 2019; Fellman et al. 2020; Ritakallio et al. 2021). The huge task-specific improvement in WM tasks led to the suggestion that training facilitates the discovery of new effective cognitive routines (Gathercole et al. 2019; Norris et al. 2019), which are useful only for the trained task and very similar tasks. Yet, the neuro-cognitive mechanisms, namely the specific cognitive strategies and their underlying neural mechanisms, have not been studied.

Here, we aimed to decipher the neurocognitive processes which underlie the gradual switch from naive, capacity-limited performance in WM tasks to successfully trained performance that is not limited by the WM bottleneck. Behaviorally, we asked what strategy underlies the improvement of very-successful learners when trained with a particularly challenging WM task. Neurally, we asked whether successful training is associated with a unique change in the pattern of brain activation.

Received: February 19, 2022. Revised: April 13, 2022. Accepted: April 14, 2022

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

For this purpose, we designed a novel, particularly challenging auditory WM task, which we termed as tone reordering (TRO). Each trial begins with a series of randomly chosen tones, which is followed by a visual presentation that specifies how these tones should be manipulated (reordered). A sequence with reordered tones is then presented and the participant must indicate if the new order corresponds to the intended one. We used a challenging perceptual task as a control task. Another group of participants was trained with pitch discrimination between tone sequences with small pitch intervals (micromelodies [MMs], see Zatorre et al. 2012). Retention and comparison of these short melodies were perceptually challenging but did not require any manipulation of the stimuli. Both groups performed both tasks during scanning before and after 40 sessions of behavioral training on their designated task.

To decipher the training-induced mechanism that allows performance to be “freed” from WM capacity limitation, we analyzed the relationship between the magnitude of behavioral learning on the WM task and its relation to the pattern of errors. We also compared brain activation pre- and posttraining for each of the tasks and as a function of training success. This comparison allowed us to further ask whether successful learning is associated with a unique change in the pattern of dorsal versus ventral activations depending on the nature of the task.

Methods

Participants

Thirty-two participants began the training study; 4 dropped out before completion, and their data were not included in the analyses. Twenty-eight participants completed the study, 14 in each training group. All participants were right-handed, native Hebrew speakers. They reported no learning disabilities, attention deficit disorders, neurological disorders, or hearing deficits and had <2 years of musical training. Participants received a standard monetary compensation for their participation in the experiment plus bonuses as described below. Ethical approval was granted by the Ethics Committee of the Hebrew University of Jerusalem. Participants were recruited using ads at the Hebrew University and via social media. Table 1 reports the demographic characteristics of the participants who completed the training.

Training tasks

WM task—TRO

As illustrated in Fig. 1A, in each trial, the participant first listened to a sequence of 3–8,400 ms tones (S1). The tones were randomly selected from a broad frequency range (250–1,000 Hz, with the limitation that the frequency of all the tones in the same sequence will be at least 20% different one from another) so that participants will not be familiarized with the stimuli during training. After

Table 1. Demographic information of the participants; mean (SD).

Group	Trained on Tone-reorder	Trained on MM
N	14	14
Female	6	7
Mean age (years)	26.3 (4.3)	25.7 (2.9)
Musical training (years)	0.32 (0.42)	0.43 (0.55)

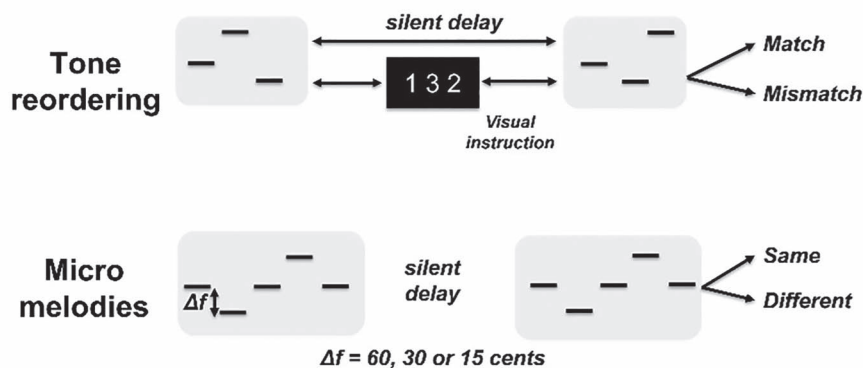
the onset of the auditory sequence, the participant had 5 s to listen and memorize it. Then, a visual string—a sequence of digits that indicated the expected reordering of the tones—was presented in the middle of the screen. After 5 more seconds, the same tones were played again (S2) either in the order specified by the visual string or in a different order. The participant’s task was to determine whether the second sequence of tones matched the permutation specified in the visual string (“match”) or not (“mismatch”). The visual instruction remained on the screen until the participant responded. In each trial, participants received visual feedback on whether or not their answer was correct. In half of the trials, S2 matched the visual instruction; in the other half, S2 did not match the visual instruction and was a random permutation of S1. In each trial, the participant had up to 10 s (from the termination of S2) to answer. Total trial length was ~15 s. This design allowed sufficient amount of time for online TRO in each trial. An interval of ~1.5 s (with up to 1-s jitter) separated between consecutive trials. Each training session consisted of 4 blocks of 28 trials each.

In the assessments in the scanner, both pre- and post-training of all sequences were composed of 3 tones. This fixed structure allowed us to compare post- to pre-training performance, and the trained to the untrained groups, under the same protocol.

Perceptual task—MMs

This task was adapted from a previous study (Zatorre et al. 2012) and was used as a perceptual learning control task that requires retention of tone sequences and fine pitch resolution but does not require any manipulation of information in WM. Following the terminology of Zatorre et al. (2012), we define MM as a melody with a constant pitch interval (frequency ratio) < 100 cents between each pair of consecutive notes (the cent scale is used to represent logarithmic frequency differences; 100 cents correspond in musical terminology to a semitone, i.e. ~6% change in frequency). Each MM consisted of 6–8 pure tones of 200 ms each, with an intertone interval of 150 ms. Thus, the total length of each MM was between 2 and 2.7 s (the 2 MMs in a trial were always the same length). The middle tone (i.e. number 3 or 4) of each MM was set to the frequency of 250 Hz. To create enough variability, there were either 2 or 3 inversions of melodic contour (changes in pitch direction) with respect to the fixed middle tone (e.g. down-down-down-**up-down-down-up** would contain three inversions, denoted in bold) in each tune.

A. Auditory Tasks



B. Protocol

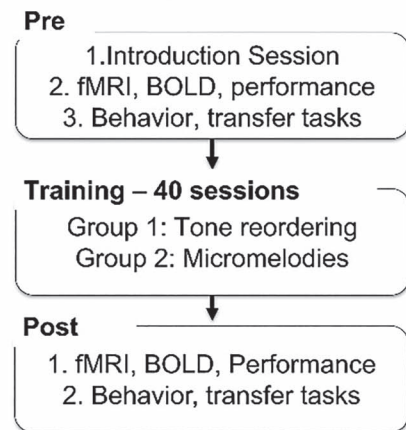


Fig. 1. Description of the training tasks and overall protocol. A) A schematic illustration of the 2 tasks. Top: Timeline of a trial in the TRO WM task as administered in the magnet pre- and posttraining (nonadaptive version). Participants listened to a sequence of 3 tones (sequence length in the adaptive task practiced by participants increased up to 8 tones). A visual string displaying the expected order of tones in the second sequence was presented 5 s after the onset of the first sequence until the end of the trial. After 5 additional seconds, participants heard another 3-tone sequence, composed of the same tones, and had to determine whether the order of its tones matches the order in the visual string (the illustration is a match). Bottom: Timeline of a trial of the perceptual MM task. Pitch interval between consecutive tones (15, 30, or 60 cents) was constant within a melody. In each trial, 2 MM were presented with a 1-s silent interval between them, and the participant had to determine whether they were identical. B) Overall testing and training protocol: 3 pretraining sessions, 40 training sessions, and 2 posttraining sessions. Pre- and posttraining sessions were identical for the 2 groups.

As shown in Fig. 1A, in each trial of the MMs task, participants were presented with pairs of stimuli (MMs) and were asked to indicate whether they were same or different. The silent delay between the two melodies was 1 s. Participants received visual feedback on whether or not their answer was correct for each trial. Total trial length was ~ 5 s. During pre- and posttraining assessments, 3 frequency intervals were used: 60, 30, and 15 cents.

Overall protocol

Initial screening of participants

All the potential participants filled a short questionnaire and were interviewed via phone in order to ensure only minimal musical background, Hebrew as mother tongue, no neurological or psychiatric disabilities, and compatibility with magnetic resonance imaging (MRI) recording.

3 pretraining sessions

Participants were invited for a baseline session. All participants understood the task and performed it above chance level. During the first session, participants were introduced to the 2 training tasks: 35 min of the TRO task (84 trials) with sequences of 3 tones, divided into 6 blocks, and then 35 min of the MMs task (108 trials) with pitch intervals of 60, 30, and 15 cents, divided into 6 blocks (pitch interval was constant within each block). In addition, they filled an MRI safety form.

In the second (or third) session, they were tested with a battery of tasks aimed to assess their overall performance in the 2 domains of subsequent training: WM and pitch perception (see details below).

In the third (or second) session, they were scanned with a Skyra 3 MRI scanner while performing each of the training tasks. All participants were scanned while

performing both tasks, though they were subsequently trained only with 1 task. Before entering the scanner, participants performed a 10-min practice refresh period of each task in order to get familiar with the MRI version of it since there were minor adaptations of each task due to scanning requirements (described in the “Scanning sessions” section). The order of sessions “2” and “3” was counterbalanced across participants.

Following the first pretraining session, each participant was randomly allocated to 1 of the 2 training groups while keeping the mean age and musical background matched between groups. After the 3 pretraining sessions, each participant was trained with 1 of the tasks, as described below. After completing the training, all participants were administered 2 posttraining sessions, which were identical to the second and third pretraining sessions: 1 dedicated to behavioral assessments and 1 dedicated to functional magnetic resonance imaging (fMRI) scans during task performance. Posttraining sessions were conducted at least 1 day, and no more than 10 days, after the end of the training period in the same order pretraining sessions were conducted.

Training protocol

Each participant trained at home for 40 sessions (5 sessions a week, 2 months) of ~ 40 min each via a designated internet site based on the Amazon AWS platform coded with JavaScript. Participants were instructed to perform the training on a computer in a quiet room at home using headphones at a comfortable and constant sound level. Participants were instructed to neutralize any computer software that could have distracting pop-up messages before beginning a training session. An experimenter monitored the frequency of training and the progress made by each participant.

Adaptive training protocols

Training of each task began with a relatively easy condition. The level of difficulty increased with the participant's progress, as follows: When overall accuracy in 2 consecutive blocks was >85%, difficulty level increased; when it was <65%, the difficulty level decreased; otherwise, the difficulty level did not change. For the TRO task, the initial difficulty level (and also at posttraining, to compare the same conditions pretraining) was sequences of 3 tones each. When a participant mastered a level (>85% accuracy), the sequences became 1 tone longer. The maximal length (and most difficult level) was 8 tones. For the MM task, difficulty level increased by decreasing pitch intervals from the introductory 150 cents to 60, 40, 30, 20, 15, 10, and 5 cents. Any given MM had 1 constant pitch interval. On half of the trials the 2 MMs were the same; in the other half, they differed. On "different" trials, the 2 MMs were of the same length and were matched for interval scale (e.g. both consisted of same size intervals), but the second MM was randomly selected from a pool of MMs such that it had a different melodic contour than the first one (at least 1 note differed between the 2 MMs).

Participants trained on either task were given feedback after each trial. To increase their motivation, they were also given a monetary bonus of 15% for each block that was performed above 65% correct and 30% for each block that was performed above 85%. At the end of each block, they received written feedback about their performance, including the percentage of earned bonus.

Behavioral assessments administered before and after training

WM tasks

Very near to near tasks

1. Noises reorder. Very near transfer effect was evaluated using the same TRO task with a different type of stimuli. Instead of pure tones, each auditory sequence consisted of 3 unfamiliar nonvocal sounds ("noises") of similar overall energy (RMS), with the same duration of 400 ms per item. The task consisted of 2 blocks of 24 trials each (50% match trials and 50% mismatch trials). Note that there were only manipulation trials in this transfer task.

2. Adaptive auditory nback. Near transfer was assessed using a different, though somewhat similar, nback WM task. In this version of nback, participants were presented with different tones and were requested to press a button whenever the newly presented stimulus was identical to the stimulus presented n steps back. We used 8 different auditory stimuli: 4 very different frequencies (100, 300, 900, and 2,700 Hz pure tones) \times 2 very different durations (130 and 350 ms). Interstimulus interval (ISI) was 1,500 ms. Overall, participants performed 15 blocks of this task. Data were analyzed only from the last 10 blocks. The first block was always of $n=2$, and the n of each of the next blocks was determined according to the participant's performance on the last block, following the same adaptive protocol as in the trained tasks: If

accuracy was >85%, n increased by 1; if accuracy was <65%, n decreased by 1; otherwise, n did not change.

Intermediate to far tasks

1. Digit span. Phonological WM skills were assessed with the Digit Span test (WAIS-III) using both subtests: "forward"—repeating lists of digit sequences in their original order (series of 2 2-digit to 9-digit sequences), and "backward"—repeating sequences in reverse order (series of 2 2-digit to 8-digit sequences). Administration and scoring followed the Wechsler manual (Wechsler et al. 1997). Items were read by a native Hebrew speaker and were presented binaurally at a comfortable sound level in a quiet room.

2. Operation span. We used the automated Operation Span test as a measure of WM capacity (Unsworth et al. 2005). In this test, participants are asked to solve a simple arithmetic problem (e.g. $(1*2) + 1 = ?$). After pressing a key to indicate they solved the problem, an answer is presented on the screen, and they are asked whether it is correct (participants are asked to ensure >85% in this task). Thereafter, an English letter appears, and participants are asked to remember the letter. On subsequent trials, after each arithmetic problem, participants are asked to remember each letter they saw in the exact order. The total score is the sum of all recalled letters in their right positions. Sequences varied between 3 and 7 trials of problem + letter. The computer program for this task can be found here: <http://englelab.gatech.edu/tasks.html>.

Auditory perception—pitch discrimination tasks

Very near to near tasks

1. MMs transpose. Very near transfer effect was evaluated using the same task as the MM training task but with transposed stimuli. Each trial was composed of a MM centered around 250 Hz and was composed of a second MM with the same pitch interval and length centered around 1,150 Hz. Participants had to decide whether the melodic contour (all pitch intervals) was the same or not (though the actual frequencies differed). A full description of the task is available in Zatorre et al. 2012.

2. 2-tone serial frequency discrimination (FD). Three protocols were administered (see Jakoby et al. 2019). In all 3, each trial was composed of 2 serially presented 50 ms tones, with an ISI of 950 ms and an intertrial interval of 1 s, which began immediately after participant's button press, indicating which tone had a higher pitch (first or second). Initial step size was 4.5%, and after 4 reversals, it was decreased to 2, 1, 0.5, and finally to 0.1%. The 3 protocols differed in their crosstrial frequency regularity: "reference low" (adaptive, 80 trials), "mixed reference" (non-adaptive, 200 trials), and "no reference" (nonadaptive, 200 trials). In the "reference low" protocol, a tone of 1,000 Hz was the lowest in each trial; the other tone was either before or after the low tone and was selected according to an adaptive 3-down, 1-up staircase protocol (Levitt 1971). In the "mixed-reference" protocol, a 1,000 Hz tone was

presented in either the first or the second interval of each trial. In the “no reference” protocol, there was no reference tone and frequency distribution was broader. In this protocol, the first frequency was selected randomly from the range of 600–1,400 Hz and the second tone was either higher or lower, following the adaptive staircase procedure described above.

Intermediate to far tasks

1. Memory for sequences of tones. In each trial, participants listened to 2 5-tone sequences and were instructed to indicate whether the 2 sequences were the same or different (see [Jakoby et al. 2019](#)). The ISI was 1,300 ms between the 2 sequences and was 275 ms between the onset of each of the 5 tones within a sequence. In “different” trials, the 2 melodies differed in the pitch of 1 tone at a varying position in each trial. There were 63 trials (32 “same” and 31 “different”), which were delivered in a fixed order.

2. Memory for melody. Participants heard 2 unfamiliar melodies in the Western major scale and were instructed to indicate whether the 2 melodies were the same or different (see [Foster and Zatorre 2010](#)). The ISI was 1,300 ms between the 2 melodies and 275 ms between each of the tones within a melody. Each melody consisted of 5–13 notes between C4 and E6 and was played with harmonic tones that were low-pass-filtered <8 kHz. All notes were 320 ms long. On half of the trials, the pitch of a single note anywhere in the melody was changed by up to ± 5 semitones (median of 2 semitones). The change maintained the key of the melody as well as the melodic contour (the order of upward and downward pitch movement in a melody without regard to magnitude). There were 60 trials, which were delivered in a fixed order.

Scanning sessions

In order to remind the tasks to the participants and let them experience the specific version of the task that was designed for the scanning session, each participant performed 10 min of each training task right before being scanned. This prescanning period included 32 trials of the TRO task (half “match” and half “mismatch”) and 45 trials (15 trials of each pitch interval size) of the MM task. For the scanning portion, a Skyra 3 MRI scanner was used. Presentation software (Neurobehavioral systems, Albany, CA, United States) was used to run the experiment and to record participants’ answers. Stimuli were presented via MRI-compatible insert earphones (Sensimetrics S14). The level of sound presentation was set to 65 dB SPL for all participants.

The entire scanning session lasted ~ 75 min per participant and consisted of several parts: T1 MPRAGE acquisition (4 min); resting-state fMRI (7 min); active fMRI recording: TRO task fMRI (14 min \times 2 blocks) + MM task (12 min \times 2 blocks)—half of the participants performed the tasks in the following order: TRO-MM-TRO-MM, and half performed in a switched order: MM-TRO-MM-TRO (for each participant, the order of the blocks from the

pretraining was kept the same for the posttraining session); diffusion MRI acquisition (7 min). The results of the resting-state fMRI and diffusion MRI are not in the scope of this paper, and therefore, we will not elaborate on these protocols.

During all fMRI data acquisition, participants were asked to keep their eyes open. The parts were separated by 2–3-min breaks. Participants were informed about the task before each run.

For each trial, participants indicated their answers after the end of the second auditory sequence by pressing 1 of 2 keys of a response device with their right hand. They had 4,100 ms (in the TRO task) or 2,000 ms (in the MM task) to respond before the next trial, which occurred between 500 and 1,000 ms after the end of the trial. For both tasks, trials within each block were distributed in a pseudorandom order, with the constraint that the same trial type (same or different) could not be repeated > 3 times in a row. In each TRO block, there were 42 trials (overall 84 trials). One third of the trials in each block (14 trials) were “retention-only” trials, with the visual instruction of 1 2 3. They served as a control condition, which requires retention only, in order to evaluate contribution of the manipulation itself to brain activations. In addition, 16 silence trials were added and were randomly distributed to serve as an implicit baseline. In each MM block, there were 54 trials (overall 108 trials) with MMs of 3 different interval sizes: 60, 30, and 15c (18 trials of each interval per block) and 24 silence trials that were added and were randomly distributed to serve as an implicit baseline. The different interval sizes were mixed in each block.

fMRI design and acquisition parameters

At the beginning of the MRI session, a high-resolution 3D anatomical MPRAGE T1-weighted image was acquired for each participant using a gradient-echo sequence (160 sagittal slices; time repetition (TR), 2300 ms; time echo (TE) 2.98 ms; flip angle (FA), 9°; matrix size, 256 \times 256; field of view (FOV), 256 \times 256 mm²; voxel size, 1 \times 1 \times 1 mm³). A gradient-echo EPI pulse sequence was used to measure whole-brain blood oxygenation level-dependent (BOLD) signal for all of the functional scans—resting-state and task blocks (42 axial slices with multiband; acceleration factor 3; TR, 1,000 ms; TE, 30 ms; FA, 62°; 3-mm slice thickness; no gap; matrix size, 64 \times 64; FOV, 192 \times 162 mm²; voxel size, 3 \times 3 \times 3 mm³). We used an event related paradigm with continuous acquisition after piloting the experiment with sparse and continuous acquisition protocols. Since both methods yielded very similar results, we opted for more data per participant and chose continuous acquisition. In each of the TRO blocks, there were 885 volumes, and in each of the MM blocks, there were 720 volumes.

Preprocessing

All image preprocessing was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, <http://www>.

fil.ion.ucl.ac.uk/spm/, London, United Kingdom). Before preprocessing, all images were checked for artifacts and were automatically aligned so that the origin of the coordinate system was located at the anterior commissure. Preprocessing included the realignment of functional images and the coregistration of functional and anatomical data. We then performed a spatial normalization (voxel size: $2 \times 2 \times 2$) of the T1 and the EPI images to the Montreal Neurological Institute (MNI) templates provided with SPM12 (MNI T1 template and EPI template, respectively). Finally, functional images were spatially smoothed (Gaussian kernel, 8 mm full-width at half-maximum).

fMRI analyses

Individual contrast maps were first calculated for each participant. A hemodynamic response function (HRF) was chosen to model the BOLD response (microtime resolution of 16 ms; microtime onset, 1; high-pass filter, 128 s). At the first level, for each participant, changes in brain regional responses were estimated by a general linear model (Friston et al. 1995). We used an event-related design: for the TRO task, each event occurring at the onset of the visual instruction (indicating the order of the to-be-expected second tone-sequence); for MM, the events occurred at the beginning of the first sound of a pair. First-level contrasts were computed separately for pre- and posttraining sessions. We then analyzed within and between group effects at the second level. Statistical inferences were performed at a threshold of $P < 0.05$ FWE cluster-corrected.

In order to identify training-related plasticity, we trained classifiers to discriminate pre- and posttraining sessions for each task on brain imaging data. As mentioned above, BOLD responses per condition (TRO: manipulation, MM: sound perception) were modeled using an HRF (same parameters as above) to generate beta maps for each participant, condition, run, and session. These maps were then used for the classification analysis. Multivariate analyses were performed using the Decoding Toolbox (Hebart et al. 2015) and LibSVM's linear support vector machine (SVM) implementation (www.csie.ntu.edu.tw/~cjlin/libsvm/). We used motion-corrected, coregistered images (but not normalized or smoothed) as input to the classifier. All classification analyses were performed using a leave-1-run-out crossvalidation procedure and a searchlight procedure, whereby the classification algorithm considers only voxels from a small sphere of space (radius = 4 voxels). Results were expressed as accuracy minus chance of category identification which was calculated using an average of the crossvalidation folds, and this value was assigned to the center voxel of the sphere. This procedure was repeated using every brain voxel as a searchlight center (~35,000–45,000 spheres), yielding local accuracy maps for the entire brain (see Albouy, Peretz, et al. 2019; Albouy et al. 2020 for similar procedure).

The analysis output was a unique map for each participant containing the classification accuracy for each voxel. For the TRO, we trained and tested the classifier to categorize pre- and posttraining sessions using manipulation trials. For the MM, we trained and tested the classifier to categorize pre- and posttraining sessions during performance of the MMs (regardless of the pitch intervals). Decoding accuracy maps were then normalized in the MNI space, smoothed with the same parameters as above, and analyzed in the second-level analysis with SPM12. Group comparisons and correlations with behavioral performance were then performed and statistical significance was established at $P < 0.05$ cluster-corrected.

Results

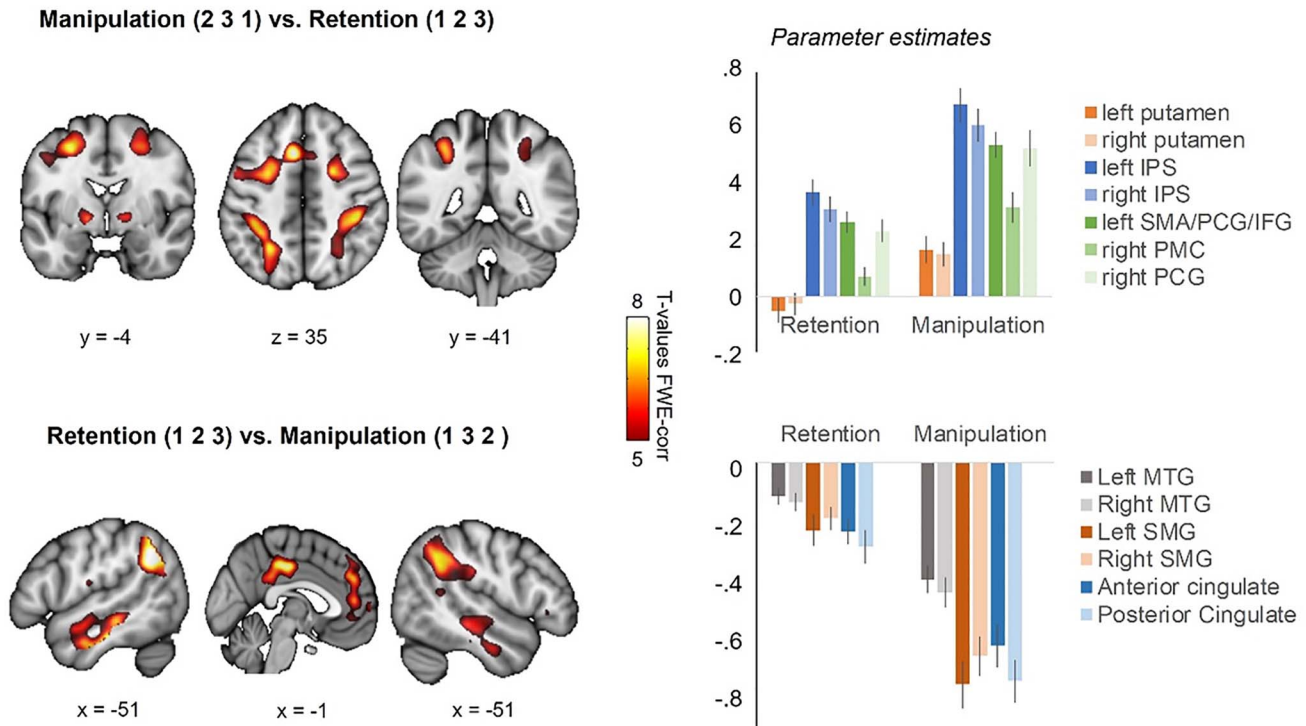
Pretraining performance and cortical activation WM task—TRO

Prior to training, the 2 groups had similar mean levels of performance in both tasks (TRO task: 63.8%; standard deviation [SD]: 10.5) and 62.5% (SD: 10.1) for the TRO and MM training groups, respectively, $P = 0.41$, 2-tailed Mann-Whitney U test). Participants were then invited to a scanning session, where they performed the task again (84 trials for WM task, now divided into 2 blocks). Mean accuracy during scanning was again similar in both groups. (67.1%; SD: 10.9), and 66.6% (SD: 10.6) for the TRO and MM training groups, respectively, $P = 0.78$, 2-tailed Mann-Whitney U test). Performance of both groups tended to be slightly better than during the baseline session, but this difference was not significant ($P = 0.26$). Each of the TRO trials required manipulation, specified by the visual string, except the visual string “1 2 3,” which means that the subsequent sequence should have the same order as the one just presented. Thus, only retention is required. These trials were used as retention-only control data. Indeed, participants showed higher accuracy (89.1%; SD: 10.5) for these retention trials than for manipulation trials (66.9%; SD: 11.4, $P < 0.01$).

For fMRI data, we expected that the TRO task will activate fronto-parietal WM regions in the dorsal stream when manipulation is required (compared to retention-only). When comparing the retention trials to manipulation (the opposite contrast), we expected a larger activation in auditory regions in the temporal lobe (Albouy et al. 2013; Kumar et al. 2016; Albouy, Caclin, et al. 2019). Figure 2A shows these 2 contrasts, which were calculated for all 28 participants in the first scanning session. Indeed, they yielded the expected patterns. The top panel shows the pattern of cortical activity during the manipulation trials compared to the retention-only trials. This contrast reveals the clear signature of a WM manipulation task—strong dorsal fronto-parietal activations in both hemispheres (see Table 2 for statistics and details).

The bottom panel shows the reverse contrast: pattern of cortical activity in the retention-only trials (“1 2 3”) compared to the manipulation trials. This

A. fMRI Pre-training: Tone reordering



B. fMRI Pre-training: Micromelodies

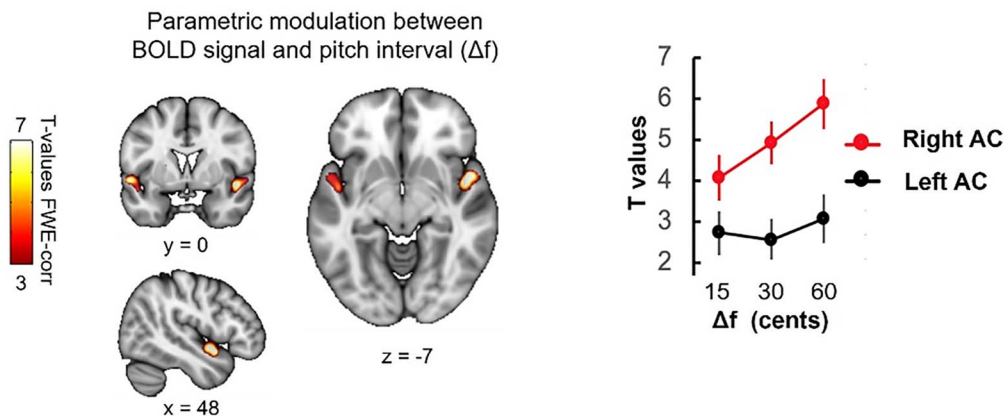


Fig. 2. Pretraining performance and pattern of brain activation of all participants ($n=28$) in both tasks (plotted on an MNI surface provided by SPM 12). **A)** WM-TRO task. Top panel: Contrast between manipulation and retention trials reveals activation in the expected dorsal fronto-parietal areas. $P < 0.05$ FWE-corrected. Right: Parameter estimates for this contrast. An increase in activations in the manipulation condition can be seen in all relevant regions: left and right putamen, left intra parietal sulcus, left supplementary motor area/pre central gyrus/inferior frontal gyrus, and right premotor cortex. Bottom panel: Contrast between retention and manipulation trials reveals activations in AC areas, supramarginal and anterior, and posterior cingulate regions. Right: Parameter estimates for this contrast. A decrease in activations (or more deactivation) in the manipulation condition can be seen in all relevant regions: anterior and posterior cingulate, left and right supramarginal gyrus, and left and right middle temporal gyri. **B)** Perceptual MM task. Left panel: Parametric modulation between pitch interval sizes (15, 30, 60 cents). BOLD signal is manifested bilaterally in activation of the lateral portion of Heschl's gyri, extending onto the superior temporal gyrus, $P < 0.05$ FWE-corrected. Right panel: Line plot of T-values for each pitch interval in the right (red line) and left (black line) ACs (regions defined by the parametric modulation). Error bars represent SEM. Note that pitch sensitivity is observed in the right but not in the left hemisphere.

contrast shows a clear pattern of bilateral activations in middle-temporal, supramarginal, and anterior and posterior cingulate gyri (see Table 2 for details). Importantly, unlike in the dorsal stream, the increased signal in this

contrast reflects reduced deactivation. Namely, as shown in the right histogram (Fig. 2A), the signal itself is negative compared to the implicit baseline in all these ventral regions in both conditions. However, it is more

Table 2. Pretraining whole-brain statistics for both tasks, all participants ($n = 28$).

Contrast	H	Region	Peak coordinates	Peak T-value	Cluster extent (k_E)	
TRO: manipulation versus retention	L	Supplementary motor area	-6 16 48	6.50	2,966	
	L	Precentral gyrus	-48 8 28	6.14		
	L	Intraparietal sulcus	-26 -66 -4	5.68	1,725	
	L	Putamen, pallidum	-14 4 -4	5.26	532	
	R	Intraparietal sulcus	44 -32 46	5.49	1,343	
			Postcentral gyrus			
	R	Precentral gyrus	50 8 24	5.66	288	
TRO: retention versus manipulation	R	Putamen, pallidum	12 2 -2	5.02	385	
	L	Middle cingulate gyrus	12 -40 36	8.28	1,716	
	L	Anterior cingulate gyrus	-12 40 48	7.56	3,652	
	R	Middle temporal gyrus, superior temporal sulcus	56 -26 -24	7.42	921	
	R	Supramarginal gyrus	58 -46 38	8.14	1,567	
	L	Middle temporal gyrus, superior temporal sulcus	-52 -24 22	7.81	861	
	L	Supramarginal gyrus	-48 -54 34	9.35	910	
MM parametric modulation with pitch interval size	R	Area 6 anterior	22 2 48	5.35	603	
	L	Planum polare	-42 -2 -6	7.51	714	
	R	Planum polare	50 6 -10	6.36	1,122	

negative in the manipulation condition (see Fig. 2A) and is hence relatively enhanced in the retention. This pattern is in line with previous studies (Todd et al. 2005; Majerus et al. 2012), which showed that enhanced dorsal recruitment (by increasing WM load) is associated with deactivation of ventral regions.

Pitch discrimination—MM task

The MM task aimed to test and train auditory retention and discrimination of short melodies with small pitch intervals. The 2 subgroups, subsequently trained with the 2 different tasks, had similar performance levels during the baseline session (mean accuracy of 63.5% [SD: 8.6] and 66.6% [SD: 9.7] for the TRO and MM training groups, respectively, $P = 0.62$, 2-tailed Mann-Whitney U test). Participants also performed the task in the subsequent scanning session (108 trials, divided into 2 blocks, a total of 36 trials of each of the 3 pitch intervals, 15, 30, and 60 cents; 100 cents = 1 semitone). Their mean accuracy in the scanner was similar (64.17% [SD: 6.7] and 65.5% [SD: 9.8] for the TRO and MM groups, respectively, $P = 0.38$, 2-tailed Mann-Whitney U test).

Figure 2B and Table 2 show the parametric modulation between the BOLD signal and the magnitude of the pitch interval in the trial, for all 28 participants. This analysis reveals the clear relationship between pitch interval and bilateral activation in the lateral portion of Heschl's Gyri, extending onto the superior temporal gyrus: the larger the interval, the stronger the activation (replicating Zatorre et al. 2012). The enhanced pitch sensitivity in the lateral portion of Heschl's Gyri (Fig. 2B, right plot, see also Table 2), is much greater in the right auditory cortex (AC), which is in line with the previous study and many others that have focused on pitch-related information processing (e.g. Zatorre and Belin 2001).

Learning WM-TRO task

In the TRO task, participants' improvement was measured during training and in the scanner comparing performance before and after training. During training, the TRO task was adaptive, and improvement was measured by the number of items added to the tone sequences. More than half (8/14) of the participants showed dramatic improvement during training. On average, sequence length increased gradually from 3 to 6.2 tones (Fig. 4B, bold line). Yet, there was substantial variability across individuals (Fig. 4B, individual performance plotted in gray lines): The 8 very-successful learners reached 8-tone sequences, the longest sequence possible, where they performed the task with 82% mean accuracy (SD: 6%). The remaining 6 participants were considerably less-successful learners: 3 did not show any improvement in sequence length, 2 reached 4-tone sequences, and 1 reached 6-tone sequences (achieving slightly below mean, this participant could be considered very-successful. Importantly, changing her group allocation does not qualitatively change the statistics reported below, see "Cortical plasticity" section below).

To decipher how these 8 individuals manage to perform the seemingly impossible instruction of randomly chosen trial-specific manipulation of eight tones, we asked them to fill a questionnaire. They were asked the following questions: (i) Did you try to manipulate all items, and if not, did you try to manipulate any item according to the visual instructions? (ii) Did you try to remember the whole sequence of tones? (iii) Did you use a specific strategy, which you can explain. (iv) Did you change strategy during the training? While the exact phrasing differed between participants, they all reported that they did not memorize nor manipulate the whole

tone sequence, as the task formally requires. They all said that they searched for a distinct tone in the original sequence, tracked its serial order, and searched for its position in the subsequent tone sequence. Typically, they did not specify whether it was a “high” or “low” pitch tone, or either, perhaps because none of them was musically trained. Thus, the discovered strategy was only partially explicit, but sufficiently so, for us to decipher it even though we did not discover it ourselves before their report.

Based on these reports, we deciphered that while, initially, participants produced the correct answer when presented with the visual string by reordering the tones according to the required manipulation, successfully trained participants did not produce an answer. Instead, they only searched for the position of an extreme, easily detected (highest/lowest pitch) tone in the first auditory sequence. They noted its serial position in this sequence (in the example of Fig. 3A, it is the highest pitch, which is in the fourth position). When the visual string was presented, they searched for the digit that denotes its serial position in the tone sequence (“4” in Fig. 3A) and then found the position of this digit in the visual string (“4” is the sixth digit). Then, when the second tone sequence was presented, they searched for this tone, and asked whether it is in the required (sixth) position. This strategy requires pitch-oddball detection in the first sequence and guided search for this tone in the second tone sequence. It also requires explicit counting for determining the exact serial positions. There may be small variations, like searching specifically for the highest or lowest tone in the first tone sequence, or attending the specific position of interest in the second tone sequence. Yet, these are all very similar strategies, and none requires TRO. Importantly, this algorithm represents a switch from the manipulation and production of the required order to confirmation. Confirmation here is probabilistic since full fit of the position of a single tone does not fully guarantee that the solution is correct, though quite high probability can be achieved. When perfectly implemented, this strategy yields over 90% accuracy for 8-tone sequences since it always yields the correct answer for “match” strings and yields 7/8 correct responses for “mismatch” ones (i.e. in 1/8 cases, the extreme pitch is located correctly but not all others, see Fig. 3B, left).

To evaluate our hypothesis regarding the deciphered strategy quantitatively, we derived 2 specific tests. First, a higher fraction of false positive is expected when extreme tones are correctly reordered and others are not. Namely, participants are expected to be more sensitive to incorrect reordering when the incorrect reordering is to tones that were tracked. This was indeed the case, as shown in Fig. 3C (stat in Table 3). Participants replied correctly and incorrectly with the same probability when the wrong reordering was only of nonsalient tones but not when relatively extreme tones were reordered. Second, optimal implementation of this strategy has

the surprising prediction of higher accuracy with 8-tone sequences compared with 3-tone sequences (or at least comparable accuracy by taking into account the short-term memory load) since the probability of correctly located extreme pitch while other tones are mislocated (which yields a false “match” response with this strategy) decreases as the length of the sequence increases (see Fig. 3B). To test this prediction, we compared the performance of the 8 successful participants with 8-tone sequences during training (in the adaptive protocol) to their posttraining performance with 3-tone sequences (in the nonadaptive magnet protocol). As predicted, successful learners performed no worse with 8-tone sequences than with 3-tone sequences, with a tendency for better performance (Fig. 3D). The advantage of longer sequences cannot be understood within the instructed task and algorithm since WM load, determined by the number of required manipulations, increases with increased sequence length.

Importantly, though posttraining performance accuracy of these very-successful participants is lower with 3 versus 8 tone sequences (Fig. 3D), it is still better than their pretraining performance with 3 tones, as measured in the magnet (pre: 68.47%, SD: 12.6 vs. post: 80.07%, SD: 6.9, $P=0.04$ in a 1-tailed Mann–Whitney U test). The less-successful learners showed slightly poorer performance pretraining and showed only marginal improvement (71.40% [SD: 11.4] compared with 61.46% pretraining [SD: 7.9]; $P=0.11$ in a 1-tailed Mann–Whitney U test). The group trained with MM did not show any improvement on the TRO task posttraining (mean accuracy posttraining was 63.30% [SD: 11.2] compared with 66.6% pretraining [SD: 10.6]).

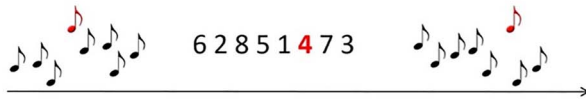
MM task

Learning the MM task was faster than learning the TRO task and was achieved within the first quarter of the training period. Still, rate of improvement varied between participants (gray lines, Fig. 5, bottom), and the 2 poorest learners reached only 40 and 60 c, respectively (Fig. 5). Posttraining accuracy level was found to be correlated with final interval size during training (Pearson’s $r=-0.57$). In the MRI scanner, the trained group showed improvement (post- vs. pretraining performance, $P=0.0009$; Fig. 5, top), while the TRO trained group showed no improvement in the MM task.

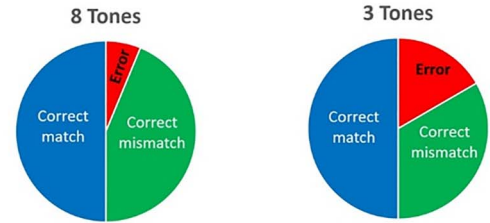
Learning specificity

As part of the testing protocol, each group performed both tasks in the magnet before and after training. Learning was task-specific, with a significant difference in the degree of improvement of the 2 groups. Namely, improvement of each group on its trained task was larger compared with the other group’s improvement in the same task. This was the case for both tasks (Mann–Whitney U test between groups on the difference postpretraining: $P=0.001$ for the MM task; $P=0.03$ for the TRO task).

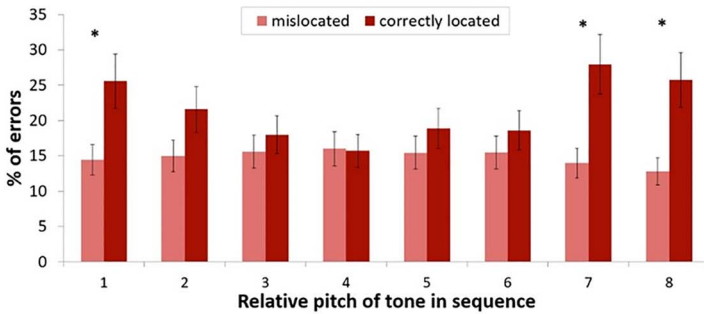
A. Strategy Illustration



B. Best-case Accuracy Rates



C. Error Rate in Mismatch Trials, 8 Tones



D. Accuracy in 3- vs. 8-tone sequences

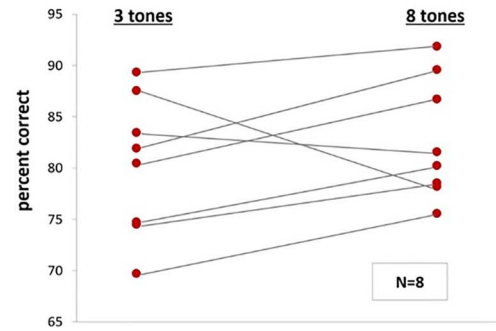


Fig. 3. Successful learners of the WM-TRO task switched their strategy—from sequence manipulation to one-tone confirmation. A) An illustration of the strategy used by the successful participants: When the first tone sequence is presented—detect the highest (or lowest) pitch tone (denoted in red) and its serial position in the tone sequence (fourth); when the visual string is presented—detect the serial position of the digit denoting said pitch’s position in the tone sequence in the visual sequence (“4” appears in the 6th position); when the second tone sequence is presented—count this number of tones (6) and assess whether this is the highest (or lowest) tone in this sequence. If so—reply “match.” B) Optimal implementation of this strategy yields higher accuracy for 8-tone compared with 3-tone sequences, since combinatorically, there is a lower chance of a mismatch with 1 prominent tone correctly located when there are more tones per sequence. C) Error rate of the eight successful learners in all mismatch trials (“false-positive” errors) of the 8-tone sequences, plotted according to the relative pitch of each tone. When extreme tones (high or low) are correctly located in the second sequence (noted as “correct” according to the extreme-pitch allocation strategy), participants tend to mistakenly say “match” significantly more than when these tones are erroneously located (see Table 3 for detailed stat). By contrast, participants’ responses are not sensitive to the position of intermediate-pitch tones. Error bars indicate SEM. Asterisks indicate significance level of $P < 0.05$ in a Wilcoxon sign-ranked test. D) Posttraining accuracy in 3-tone sequences compared to accuracy in 8-tone sequences during training (all 8-tones trials included) for the 8 participants who have reached 8-tone sequences—accuracy for 8 tones tends to be higher for 6/8 successful participants.

Table 3. Mean error rate across participants in mismatch trials (“false-positive” errors) of 8-tone sequences, presented separately for each tone, numbered according to their relative pitch in the sequence, from lowest to highest. P value is in Wilcoxon sign-ranked test between the 2 cases (location matching or not matching the visual sequence). Participants tend to have false positive when prominent tones (1, 2, 7, 8) are correctly ordered in the second sequence.

Relative pitch of the tone in the sequence	1	2	3	4	5	6	7	8
(a) Mean % errors when correctly located	25.58	21.58	17.98	15.69	18.91	18.62	27.96	25.73
SEM of (a)	3.28	3.22	2.98	3.14	3.29	3.05	3.27	3.10
(b) Mean % errors when mislocated	14.43	14.97	15.60	16.00	15.45	15.49	13.97	12.78
SEM of (b)	3.53	3.33	4.78	3.99	3.23	4.47	4.69	6.89
P for (a) versus (b)	0.031	0.016	0.08	1	0.69	0.08	0.01	0.01

Bold font indicates significant p values for the difference between (a) and (b).

Since the 2 tasks substantially differed in their WM versus perceptual requirements, we administered a series of additional tasks to assess transfer to untrained tasks, whose cognitive demands were more similar to the trained ones (“near transfer,” see Methods and Table 4). We assessed 4 additional WM tasks: Noise Reorder, which had the same structure as the TRO task but used unfamiliar complex sounds ($P=0.292$), and 3 other well-studied WM tasks: Digit Span backward ($P=0.391$), Operation Span, which requires retention

of letter sequences while solving simple arithmetic questions ($P=0.285$), and auditory nback ($P=0.32$). We conducted MANOVA on the % difference between pre- and posttraining results and found no difference between the TRO-trained group compared with the MM trained group in any of these tasks.

A similar analysis for the MM trained group found similar specificity, except for near transfer to a task with the same stimuli structure (MMs) and task demand, but the comparison sequence is presented in a different

Table 4. Mean (SD) performance of each group in each transfer task, before and after training.

Task	Domain	Group	Pretraining	Posttraining	Uncorrected P value
Noise reorder (% correct)	Working memory (WM)	TRO-trained	71.57 (17.3)	81.43 (14.09)	P = 0.215
		MM-trained	74.57 (13.8)	79.71 (14.59)	
Auditory nback (task score)		TRO-trained	634.35 (122.12)	680.45 (174.24)	P = 0.37
		MM-trained	721.55 (140.02)	781.91 (159.23)	
Digit span backward (raw score)		TRO-trained	7.21 (2.33)	8.5 (2.03)	P = 0.067
		MM-trained	8.21 (1.85)	8.29 (2.67)	
Operation span (sum of partial scores)		TRO-trained	52.43 (9.95)	57.7 (10.48)	P = 0.025
		MM-trained	63.93 (7.58)	62.21 (6.7)	
MM transpose (% correct)	Pitch perception	TRO-trained	59.623 (0.078)	60.417 (0.067)	P = 0.042
		MM-trained	64.583 (0.127)	72.917 (0.153)	
Memory for sequences of tones (% correct)		TRO-trained	66.44 (5.2)	60.77 (9.37)	P = 0.175
		MM-trained	67.01 (10)	67.91 (8)	
Memory for melody (% correct)		TRO-trained	64.76 (4.57)	64.64 (7.71)	P = 0.09
		MM-trained	66.67 (11.23)	67.74 (10.91)	
2-tone FD (% correct)		TRO-trained	65.71 (12.78)	65.23 (14.46)	P = 0.112
		MM-trained	65.13 (10.08)	67.76 (9.58)	

We used 1-tailed Wilcoxon sign-ranked test to compare pre- and posttraining performance within each group and used 1-tailed Mann–Whitney U test to compare the magnitude of improvement (posttraining score minus pretraining score) between groups, each containing 14 participants. Top, WM tasks; Bottom, pitch perception tasks. Tasks are presented from near to far, and names of very-near tasks are written in bold. Bold P values are significant (without correction for multiple comparisons). Transfer of WM training, was found only for Operation Span, but it does not “survive” correction for multiple comparisons. Given that there is no transfer to “nearer” tasks, like noise reordering, and hence transfer is not predicted, correction for multiple comparisons is required, and hence transfer is not significant. Transfer of the perceptual MM task to the very near transposed version, which is more difficult than the trained task, is significant, and needs no correction due to our a priori prediction of transfer (this observation replicates; Zatorre et al. 2012) and due to it being a very near task. No transfer was found to the any of the other test tasks.

frequency range (% improvement in MM trained group = 8.33, % improvement in TRO trained group = 0.79, $P = 0.042$; more details in Table 4). The test task requires transposition—the ability to compare frequency intervals across frequencies. This transfer replicates previous findings for this task (Zatorre et al. 2012).

Cortical plasticity

WM-TRO task

In order to detect brain plasticity associated with WM training, we first compared BOLD signal post- and pretraining for each training group while performing the TRO task. The MM training group did not show any significant difference in BOLD activity, which is in line with their lack of behavioral modifications. In the TRO training group, post- versus pretraining contrast (manipulation vs. retention first level contrast) showed clusters of increased BOLD signal in left superior temporal sulcus, left and right supramarginal gyri, and bilateral anterior and middle cingulate gyri (Fig. 4A, middle and right panels, cluster-corrected, see also Table 5). This spatial distribution is similar to that of the retention versus manipulation trials pretraining, as shown in Fig. 2A. As post hoc tests, we extracted the parameter estimates for each group and each session using Marsbar (Brett et al. 2002) in these regions (statistics of significant clusters in Table 5). A 2-way repeated-measures ANOVA with session as within-participant factor and group as between-participants factor showed a significant main effect for session ($F(1, 26) = 47.2, P < 0.001$), and more importantly, an interaction between session and group ($F(1, 26) = 13.6, P < 0.001$). Post hoc tests (Tukey-corrected) revealed that while the MM and TRO groups recruited the ventral pathway to the same extent pretraining ($P = 0.33$), the TRO group showed a larger increase of activity in

this network (during manipulation trials) compared with the MM group ($P = 0.007$). The recruited ventral regions are similar to those associated with oddball detection in both auditory and visual tasks, particularly when the oddball was task-relevant (Kim 2014). Still, the pattern of recruited ventral regions is general and is also similar to that of the default mode network (DMN, Shulman et al. 1997; Raichle et al. 2001; Buckner et al. 2008), whose activity was found to be modified by task difficulty. As in the DMN, the training-induced enhancement reflects reduced deactivation. Hence, this effect might reflect reduced task difficulty following training. This pattern, which is similar to that found pretraining in retention-manipulation (Fig. 2), may also reflect reduced load of the WM system as a consequence of discovering an efficient strategy by most participants, leading to reduced deactivation (Todd et al. 2005; Majerus et al. 2012).

To better understand the specific strategy discovered by the very-successful learners, we compared their post- and pretraining brain activities (8/14) to that of the less-successful WM learners (6/14). As shown in Fig. 4, 1/14 participants was a slightly below-average learner, and as such, could have been allocated to the successful-learners' group. Allocating him to the successful-learners' group does not qualitatively change the reported statistics (see Table 4). Note that although dividing the group yields 2 somewhat-small groups, the data set of each participant is rather large and very stable.

Post hoc tests showed no difference in the extent of ventral recruitment between the successful ($n = 8$) and less-successful learners ($n = 6$): repeated-measures ANOVA with session (pre and post) and group (successful and less-successful learners) found a main effect of

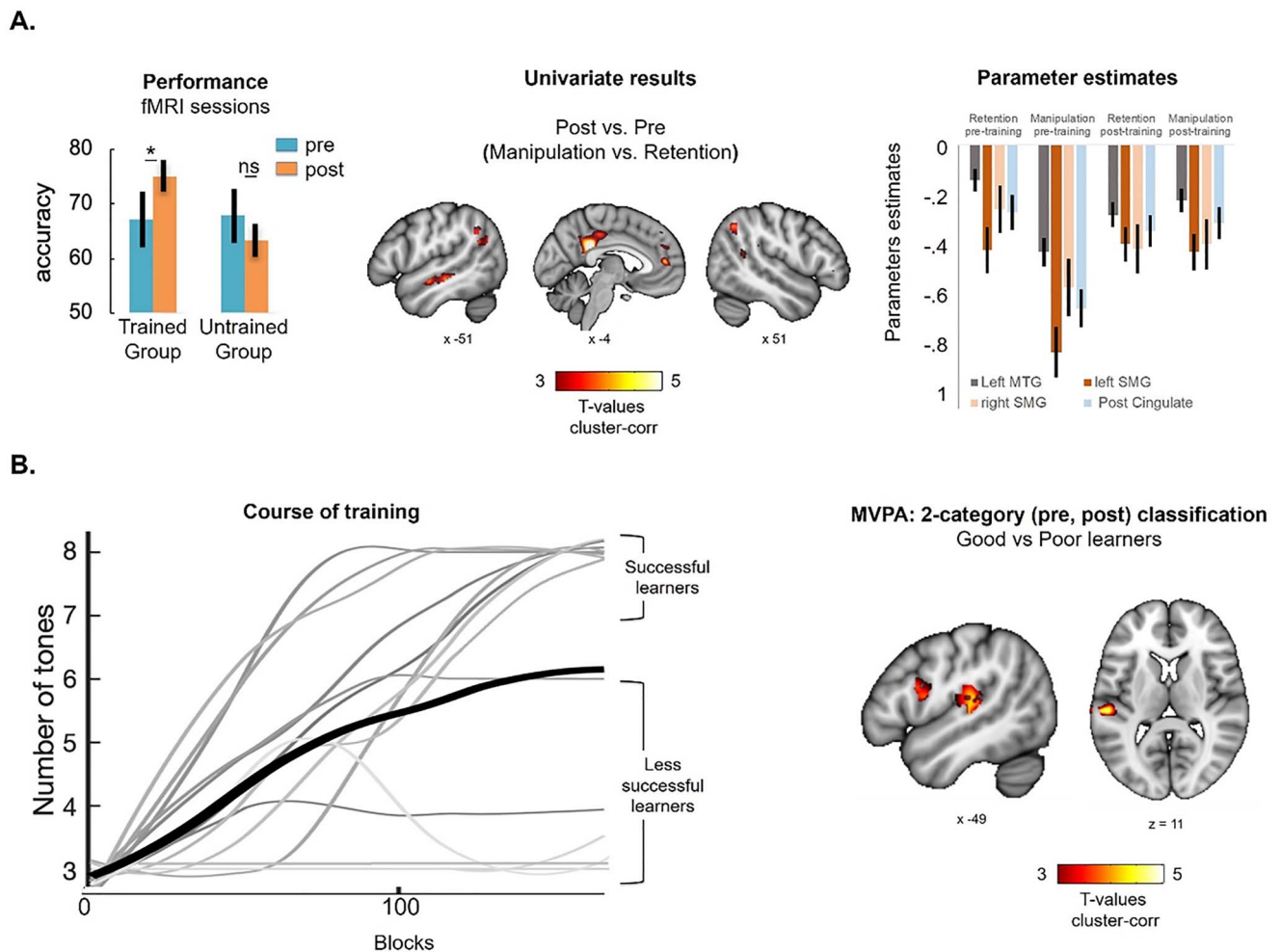


Fig. 4. Brain activity associated with training induced behavioral improvement, TRO. A) Left: Average accuracy (% correct) of each training group in the TRO task during the fMRI scans, pre- (blue) and posttraining (orange) (n of each group = 14, error bars indicate SEM). Improvement is found only in the group trained with TRO. Middle: Univariate results for trained participants ($n = 14$); contrast between post- and pretraining activity for the manipulation versus retention (first-level contrast $P < 0.05$ cluster-corrected) reveals a clear ventral activation, similar to that found pretraining in retention versus manipulation (Fig. 2, top). Results are plotted on an MNI surface provided by SPM 12. Right: Parameter estimates for this contrast. A decrease in activations (or more deactivation) in the manipulation condition can be seen in all relevant regions: posterior cingulate, left and right SMG, MTG. B) Left: Rate of improvement—the (adaptively increasing) number of tones in a sequence as a function of training session, plotted separately for each participant ($n = 14$, gray lines) and the average of all trained participants (black line). Eight participants reached 8-tone sequence level and were classified as successful learners. The other 6 were classified as less-successful learners. Each session contained 4 blocks, resulting in overall 160 blocks per participant. Right: MVPA results for the 2-class decoding (decoding pre- and postsession)—contrast between successful and less-successful learners ($P < 0.05$ cluster-corrected) shows clusters in left auditory and inferior frontal regions.

Table 5. Post- versus pretraining whole-brain statistics for the TRO and MM tasks.

Analysis	H	Region	Peak coordinates	Peak T-value	Cluster extent (k_E)
Univariate: TRO post- versus pre-WM group (manipulation vs. retention)	L	Middle temporal gyrus	58 -14 -10	5.73	994
	L	Angular, supramarginal gyrus	-46 -54 34	4.72	596
	L	Middle cingulate	-6 -42 34	5.30	1,248
	R	Middle cingulate	12 -40 34	4.62	
MVPA: TRO successful (8) versus less-successful (6) learners	R	Angular, supramarginal gyrus	56 -52 18	3.69	119
	L	Planum temporale	-56 -26 8	4.17	582
MVPA: TRO successful (9) versus less-successful (5) learners	L	Opercular part IFG	-52 10 22	3.16	191
	L	Planum temporale	-56 -26 8	3.29	754
MVPA: MM regression between decoding accuracy and training slope	L	Opercular part IFG	-56 16 26	3.52	202
	R	Superior temporal gyrus	54 -14 -8	4.16	527
	R	Inferior frontal gyrus	42 52 -6	5.96	511

session ($F(1, 12) = 33.30, P < 0.001$), but no group effect, or session by group interaction ($P = 0.57$). This suggests that the difference between successful and less-successful learners is not in the magnitude of activation of specific brain regions.

We then applied a whole-brain searchlight analyses (MVPA, SVM, leave-1-out crossvalidation procedure, cluster correction, see Methods) asking whether we could classify pre- and posttraining sessions in each participant. We performed a second-level analysis on the accuracy maps, which showed that the decoding accuracy in the left inferior frontal gyrus and left AC were significantly higher for the successful learners compared to the less-successful learners (Fig. 4B and Table 5). Namely—the pattern of activity in the left AC and left IFG in successful learners, who used the extreme-pitch allocation strategy, differed in post- versus pretraining while this was not observed in less-successful learners (who did not). These ventral regions were previously found to be involved not only in pitch perception but also in short-term memory encoding and retrieval of tonal stimuli, without manipulation (Albouy et al. 2015; Albouy, Peretz, et al. 2019). Thus, successful learners' strategy recruited the AC.

MM task

In order to detect brain plasticity associated with MM training, we compared BOLD signal post- versus pretraining for each training group while performing the MM task in a univariate analysis. At the whole-brain level, neither the MM nor the TRO training group showed any significant difference in BOLD activity between the 2 scanning sessions. However, the univariate analysis did not take into account individual differences in learning nor it is sensitive to changes in patterns of activity as opposed to magnitude. In order to identify the potential differences in brain plasticity associated with the MM training using a more sensitive technique, we applied whole-brain searchlight analyses (SVM, leave-1-out crossvalidation procedure and cluster correction, see Methods) to classify pre- and posttraining sessions of each participant. Group comparison on the decoding accuracy maps did not show any significant differences. We then correlated these decoding accuracy maps with the behavioral training slopes in the MM task and found that decoding accuracy in the right fronto-temporal network was positively correlated ($P < 0.05$) with the training slope (Fig. 5B, cluster-corrected), indicating that though no univariate effect was found, the rate of improvement is associated with the pattern of activity in the right AC and right IFG, as one would expect based on previous findings (Albouy et al. 2013; Kumar et al. 2016; Bianchi et al. 2017).

Discussion

Practice, particularly using adaptive protocols, ubiquitously improves performance in the trained task, but

what is it that we learn during training remains an open question, with huge conceptual and applied implications. In line with previous studies, we found that people can substantially improve in both a challenging perceptual auditory task and in a difficult WM task, though in both cases, learning was specific to the trained conditions. Using the perceptual MM task, we were able to replicate a greater response to pitch variation in the right AC (Zatorre et al. 2012). We did not see a change in overall BOLD magnitude post training with this task, but using a multivariate classifier, we were able to demonstrate that the pattern of activity in right temporal and frontal regions was related to learning rate, thus extending previous findings implicating this circuitry in the learning of fine-grained pitch patterns. However, the main novelty of the current study is in deciphering the neurocognitive mechanisms that can be successfully recruited when training on a challenging WM task. Since the required manipulations were determined in each trial de novo, participants could not predict them. Successful participants implicitly discovered that tracking the position of a single, easy-to-detect tone, provides sufficient information for quite successful, though not perfect, performance. This strategy is particularly tailored to the adaptive protocol, since when perfectly performed, the probability of it being correct increases with sequence length, which is in contrast to the formal task demands.

When people train with perceptual tasks, they tend to adopt a strategy, often implicitly, and keep it (Ahissar and Hochstein 2004; Ahissar et al. 2009). By contrast, in the TRO task, the WM bottleneck does not enable successful performance without discovering an efficient strategy. In the current case, participants switched from producing the required manipulation, which is limited by WM bottleneck, to performing a partial, yet “good-enough” verification. This strategy takes advantage of the specific task structure where participants are given an answer that is either correct or chosen randomly from other permutations. In this protocol, verifying the adequacy of the position of 1 tone is very informative. In order to implement such strategy, a number of different cognitive components need to be brought into play. These might include attention, encoding, and retention. Future studies will have to distinguish which components are most directly related to the changes in brain activations that we observed.

In both the WM and the perceptual tasks, training-induced plasticity was associated with modifications of activity in the temporal lobe. For the perceptual task, this observation is in line with previous reports for learning-induced increase in the accuracy of the retention of pitch sequences (Zatorre et al. 2012; Bianchi et al. 2017). Successfully trained participants in the TRO task manifest training-induced information in the auditory regions used by performers of the MM task (Planum temporale, IFG). This observation suggests their use of ventral circuitry for more accurate retention and discrimination of the tone sequences. It is worth noting

Micromelodies

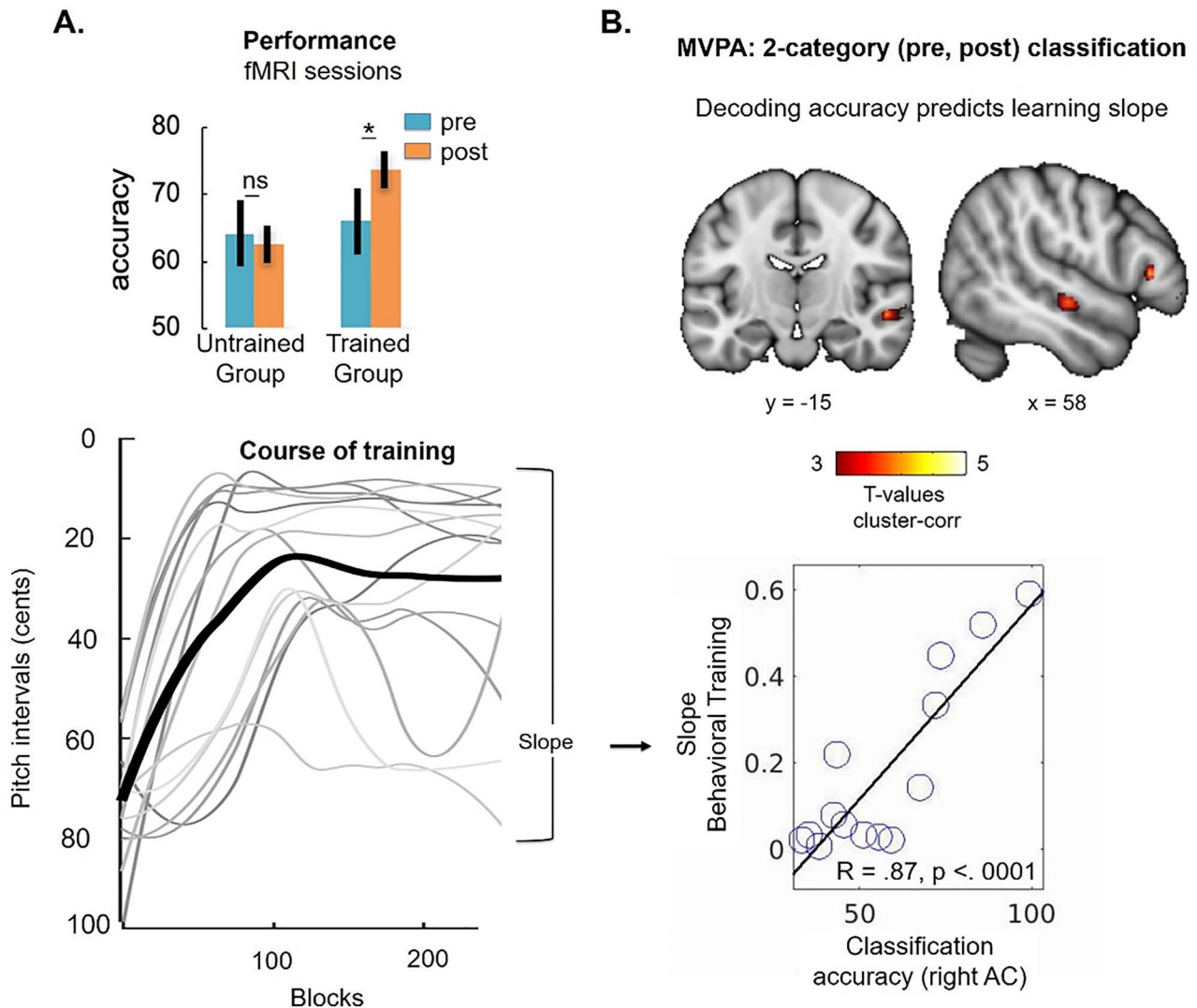


Fig. 5. Brain activity associated with training induced behavioral improvement, MM. A) Top: Average accuracy (percent correct) of each training group in the MM task shows improvement only for the MM-trained group. Bottom: Rate of improvement—the (adaptively decreasing) pitch interval (denoted in cents) as a function of session number, plotted separately for each participant ($n = 14$, gray lines) and their average (black line). Each session contained 6 blocks, resulting in overall 240 blocks per participant. B) Top: MVPA results for the 2-class classification (decoding pre- and postsession): Whole-brain regression between the behavioral slope and decoding accuracy, $P < 0.05$, cluster-corrected reveals right inferior frontal gyrus and right AC. Bottom: Scatter plot of individual learning slopes as a function of decoding accuracy in the right AC, showing that rate of improvement was correlated with changes in the pattern of activation in this region.

that the same regions are known to be activated during encoding and retrieval of tonal stimuli (Albouy et al. 2015; Albouy, Peretz, et al. 2019). Thus, while naive performers of the TRO task break the tone sequence into individual components, remember each tone separately and then manipulate them, successfully trained participants use the representation of the whole sequence, as they search for the salient pitch component within this sequence. This perceptual process of retention and search is compatible with modifications in the activity of pattern of auditory regions. Hence, such evaluation involves recruitment of the ventral streams (Ahissar et al. 2009) and does not depend on limited WM capacity (Myers et al. 2018).

Is switching to perceptual processes a general mechanism for attaining proficiency in WM tasks?

Imaging studies consistently find that novel tasks, whether perceptual (Daikhin and Ahissar 2015) or cognitive (Duncan et al. 2000), activate the frontoparietal networks. These networks seem to implement serial tasks whose strategies are not yet “hard-wired” in the system (Duncan and Owen 2000) and are therefore termed the WM network. We found that successful practice with a specific WM task leads to the use of informative perceptual traces, which reduces manipulation demands. It would be interesting to see in the future the extent to which the TRO task is representative of other

WM tasks in this aspect. We believe that reducing WM demands is a necessary part of forming efficient task-specific strategies for WM training tasks even though it can be only a part of it (see, e.g., Malinovitch et al. 2020). The crux of this process is the understanding, often implicit, of the informative aspects of the tasks so that successful performance does not require stimulus manipulations, such as swapping the position of items using limited capacity WM. In the TRO task, it is the implicit understanding that good-enough performance can be attained by tracking the position of one item.

In the well-known nback WM task, where participants are asked to detect a repetition with an interval of n items in a series of serially presented items (Kirchner 1958; Mackworth 1959; Moore and Ross 1963), trained participants often show huge task specific improvement (e.g. Jakoby et al. 2019). It was recently shown that, as in the TRO task, trained participants use a more efficient strategy (Laine et al. 2018). In the nback case, it is the implicit understanding that rather than shifting item position in WM—the attended position in WM should be shifted—reducing the number of manipulations to 1 regardless of the formal load of n . As in the TRO task, not all participants show this huge improvement, but all those that do, discovered the efficient strategy (Malinovitch et al. 2020). Thus, in both the nback and the TRO tasks, increasing manipulation load with increasingly complex samples was replaced with increasing load on retention. Though, to the best of our knowledge, these are the only 2 WM tasks for which the specific strategy was revealed, we propose that they reflect a prevalent pattern.

Successful WM training and skill acquisition

We propose that similar processes also underlie the acquisition of complex skills. The term skill relates to becoming experts in complex tasks, which involves many routines and a rich vocabulary and requires years of training. For example, becoming a master in chess takes years, yet acquiring the perceptual subcomponents, such as the opening arrangement of the chess pieces, is quickly acquired, suggesting that acquisition of complex skills reflect the accumulation of many subcomponents. These subcomponents are stored as perceptual configurations, as suggested by imaging studies (Bilalić et al. 2011). These studies find that chess masters activate visual perceptual regions that store holistic configurations (like the Fusiform Face Area) when presented with an arrangement of chess pieces.

While chess expertise is acquired by a small portion of the population, acquiring expertise in reading is almost mandatory. Reading also begins as a WM task, where readers actively “merge” subunits which they hold in accessible short-term memory. It is therefore serial and demanding. Indeed, it activates the dorsal, serial phonological route (Cohen and Dehaene 2009; Dehaene 2009). Becoming a proficient reader takes huge amounts of

practice, but this, we propose, largely stems from the very large vocabulary (of syllables and words) that is used. While beginners use serial decoding for reading unfamiliar strings, years of practice yield expert-level information-selective reading, and the gradual formation of a reading-specific region, which is often termed the visual word form area. Activity in this region is sensitive to familiar, holistic forms of presentation of letter sequences (McCandliss et al. 2003). Importantly, expert readers learn to directly access the relevant information and discard redundant information. For example, they are insensitive to the order of letters when letter transposition does not yield a valid alternative, while they are sensitive to such transpositions in writing systems that are less redundant (Velan and Frost 2007). In reading, successful practice recruits ventral regions, and produces the specificity of the ventral visual word form area, whose activation is associated with reduced activity in fronto-parietal regions compared with novice readers (Dehaene 2009).

Conclusion

We found that following successful training, performance of a seemingly impossible WM task becomes more similar to that of a perceptual task. With training, people can implicitly find probabilistically efficient routines, tailored to the specific task design, which allow replacing limited manipulation capacity with retention of the whole stimulus sequence and searching for the adequacy of relevant information. This switch is associated with more efficient use of stimulus-specific ventral regions, which support retention of these stimuli. This process changes the mainly dorsal-based processes, which allow separate access to each item (and its manipulation) in active memory, yet have a very limited capacity, to recruitment of the ventral stream, including sensory regions. These regions can subserve active retention of task-informative stimuli but not their reordering of their composing components.

Acknowledgements

We thank Luba Daikhin for her help in early stages of this study. We also thank Neria Saada for programming the training internet site.

Funding

This work was supported by the Canadian Institutes of Health Research, the International Development Research Center, the Israeli Science Foundation, and the Azrieli Foundation (grant No. 2425/15). It was also supported by the Healthy Brains for Healthy Lives initiative of McGill University under the Canada First Research Excellence Fund. M.A. is supported by a personal grant from the Israel Science Foundation (grant No. 1650/17) and the ERC (grant No. 833694). R.J.Z. is

supported by funds from the Canadian Institutes of Advanced Research.

Conflict of interest statement: None declared.

References

- Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci*. 2004;8(10):457–464. <https://doi.org/10.1016/j.tics.2004.08.011>.
- Ahissar M, Nahum M, Nelken I, Hochstein S. Reverse hierarchies and sensory learning. *Philos Trans R Soc B Biol Sci*. 2009;364(1515):285–299. <https://doi.org/10.1098/rstb.2008.0253>.
- Albouy P, Mattout J, Bouet R, Maby E, Sanchez G, Aguera P.E., Daligault S, Delpuech C, Bertrand O, Caclin A, and Tillmann B. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*. 2013;136(5):1639–1661. <https://doi.org/10.1093/brain/awt082>.
- Albouy P, Mattout J, Sanchez G, Tillmann B, Caclin A. Altered retrieval of melodic information in congenital amusia: insights from dynamic causal modeling of MEG data. *Front Hum Neurosci*. 2015;9(FEB):1–13. <https://doi.org/10.3389/fnhum.2015.00020>.
- Albouy P, Weiss A, Baillet S, Zatorre RJ. Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron*. 2017;94(1):193–206.e5. <https://doi.org/10.1016/j.neuron.2017.03.015>.
- Albouy P, Caclin A, Norman-Haignere SV, Lévêque Y, Peretz I, Tillmann B, Zatorre RJ. Decoding task-related functional brain imaging data to identify developmental disorders: the case of congenital amusia. *Front Neurosci*. 2019;13(OCT):1–13. <https://doi.org/10.3389/fnins.2019.01165>.
- Albouy P, Peretz I, Bermudez P, Zatorre RJ, Tillmann B, Caclin A. Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia. *Hum Brain Mapp*. 2019;40(3):855–867. <https://doi.org/10.1002/hbm.24416>.
- Albouy P, Benjamin L, Morillon B, Zatorre RJ. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*. 2020;367(6481):1043–1047. <https://doi.org/10.1126/science.aaz3468>.
- Bayliss DM, Jarrold C, Baddeley AD, Gunn DM. The relationship between short-term memory and working memory: Complex span made simple? *Memory*. 2005;13(3–4):414–421. <https://doi.org/10.1080/09658210344000332>.
- Bianchi F, Hjortkjær J, Siebner HR, Dau T. *NeuroImage Subcortical and cortical correlates of pitch discrimination: evidence for two levels of neuroplasticity in musicians*. *Neuroimage*. 2017;163(July):398–412. <https://doi.org/10.1016/j.neuroimage.2017.07.057>.
- Bilalić M, Langner R, Ulrich R, Grodd W. Many faces of expertise: fusiform face area in chess experts and novices. *J Neurosci*. 2011;31(28):10206–10214. <https://doi.org/10.1523/JNEUROSCI.5727-10.2011>.
- Brett M, Anton J-L, Valabregue R, & Poline J-B. Region of interest analysis using an SPM toolbox. In: *8th International Conference on Functional Mapping of the Human Brain*, Sendai; 2002;16(2):497.
- Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci*. 2008;1124:1–38. <https://doi.org/10.1196/annals.1440.011>.
- Champos AS, Petrides M. Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. *Proc Natl Acad Sci*. 2007;104(37):14837–14842. <https://doi.org/10.1073/pnas.0607101104>.
- Champos AS, Petrides M. Dissociation within the Frontoparietal network in verbal working memory: a parametric functional magnetic resonance imaging study. *J Neurosci*. 2010;30(10):3849–3856. <https://doi.org/10.1523/JNEUROSCI.0097-10.2010>.
- Cohen L, Dehaene S. Ventral and dorsal contributions to word reading. In: *The cognitive neurosciences*. 4th ed. Massachusetts Institute of Technology: Cambridge (MA); 2009. pp. 789–804
- Conway ARA, Kane MJ, Engle RW. Working memory capacity and its relation to general intelligence. *Trends Cogn Sci*. 2003;7(12):547–552. <https://doi.org/10.1016/j.tics.2003.10.005>.
- Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci*. 2001;24(1):87–114. <https://doi.org/10.1017/S0140525X01003922>.
- D'Esposito M, Aguirre GK, Zarahn E, Ballard D, Shin RK, Lease J. Functional MRI studies of spatial and nonspatial working memory. *Cogn Brain Res*. 1998;7(1):1–13. [https://doi.org/10.1016/S0926-6410\(98\)00004-4](https://doi.org/10.1016/S0926-6410(98)00004-4).
- Daikhin L, Ahissar M. Fast learning of simple perceptual discriminations reduces brain activation in working memory and in high-level auditory regions. *J Cogn Neurosci*. 2015;27(7):1308–1321. https://doi.org/10.1162/jocn_a_00786.
- Dehaene S. *Reading in the brain*. Penguin; 2009. [aAC]
- Duncan J, Owen AM. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci*. 2000;23(10):475–483.
- Duncan, J., Seitz, R.J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F.N. and Emslie, H. A neural basis for general intelligence. *Science*. 2000;289(5478):457–460. <https://doi.org/10.1126/science.289.5478.457>.
- Fellman D, Jylkkä J, Waris O, Soveri A, Ritakallio L, Haga S, et al. The role of strategy use in working memory training outcomes. *J Mem Lang*. 2020;110(June 2019):104064. <https://doi.org/10.1016/j.jml.2019.104064>.
- Foster NEV, Zatorre RJ. Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*. 2010;53(1):26–36. <https://doi.org/10.1016/j.neuroimage.2010.06.042>.
- Foster NEV, Halpern AR, Zatorre RJ. Common parietal activation in musical mental transformations across pitch and time. *NeuroImage*. 2013;75:27–35. <https://doi.org/10.1016/j.neuroimage.2013.02.044>.
- Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SCR, Frackowiak RSJ, Turner R. Analysis of fMRI time-series revisited. *NeuroImage*. 1995;2(1):45–53.
- Gathercole SE, Dunning DL, Holmes J, Norris D. Working memory training involves learning new skills. *J Mem Lang*. 2019;105(December 2018):19–42. <https://doi.org/10.1016/j.jml.2018.10.003>.
- Halford GS, Cowan N, Andrews G. Separating cognitive capacity from knowledge: a new hypothesis. *Trends Cogn Sci*. 2007;11(6):236–242. <https://doi.org/10.1016/j.tics.2007.04.001>.
- Hebart MN, Görgen K, Haynes J-D. The decoding toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front Neuroinform*. 2015;8:88 <https://www.frontiersin.org/article/10.3389/fninf.2014.00088>.
- Hitch GJ, Towse JN, Hutton U. What limits children's working memory span? Theoretical accounts and applications for scholastic development. *J Exp Psychol Gen*. 2001;130(2):184–198. <https://doi.org/10.1037/0096-3445.130.2.184>.
- Holmes J, Gathercole SE, Dunning DL. Adaptive training leads to sustained enhancement of poor working memory in children. *Dev Sci*. 2009;12(4):1–7. <https://doi.org/10.1111/j.1467-7687.2009.00848.x>.
- Jakoby H, Raviv O, Jaffe-Dax S, Lieder I, Ahissar M. Auditory frequency discrimination is correlated with linguistic skills, but its training does not improve them or other

- pitch discrimination tasks. *J Exp Psychol Gen.* 2019;148(11):1953. <https://doi.org/10.1037/xge0000573>.
- Kim H. Involvement of the dorsal and ventral attention networks in oddball stimulus processing: a meta-analysis. *Hum Brain Mapp.* 2014;35(5):2265–2284. <https://doi.org/10.1002/hbm.22326>.
- Kirchner WK. Age differences in short-term retention of rapidly changing information. *J Exp Psychol.* 1958;55(4):352–358.
- Klingberg T. Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cereb Cortex.* 1998;8(7):593–601. <https://doi.org/10.1093/cercor/8.7.593>.
- Klingberg T. Training and plasticity of working memory. *Trends Cogn Sci.* 2010;14(7):317–324. <https://doi.org/10.1016/j.tics.2010.05.002>.
- Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD. A brain system for auditory working memory. *J Neurosci.* 2016;36(16):4492–4505. <https://doi.org/10.1523/JNEUROSCI.4341-14.2016>.
- Laine M, Fellman D, Waris O, Nyman TJ. The early effects of external and internal strategies on working memory updating training. *Sci Rep.* 2018;8(1):1–12. <https://doi.org/10.1038/s41598-018-22396-5>.
- Levitt HCC. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am.* 1971;49(2B):467–477.
- Mackworth JF. Paced memorization in a continuous task. *J Exp Psychol.* 1959;58(3):206–211.
- Majerus S, Attout L, D'Argembeau A, Degueldre C, Fias W, Maquet P, Martinez Perez T, Stawarczyk D, Salmon E, Van der Linden M, and Phillips C. Attention supports verbal short-term memory via competition between dorsal and ventral attention networks. *Cereb Cortex.* 2012;22(5):1086–1097. <https://doi.org/10.1093/cercor/bhr174>.
- Malinovitch T, Jakoby H, Ahissar M. Training-induced improvement in working memory tasks results from switching to efficient strategies. *Psychon Bull Rev.* 2021;28(2):526–536. <https://doi.org/10.3758/s13423-020-01824-6>.
- McCandliss BD, Cohen L, Dehaene S. The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn Sci.* 2003;7(7):293–299. [https://doi.org/10.1016/S1364-6613\(03\)00134-7](https://doi.org/10.1016/S1364-6613(03)00134-7).
- Meiran N, Dreisbach G, von Bastian CC. Mechanisms of working memory training: insights from individual differences. *Intelligence.* 2019;73(February):78–87. <https://doi.org/10.1016/j.intell.2019.01.010>.
- Melby-Lervåg M, Hulme C. Is working memory training effective? A meta-analytic review. *Dev Psychol.* 2013;49(2):270–291. <https://doi.org/10.1037/a0028228>.
- Melby-Lervåg M, Redick TS, Hulme C. Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: evidence from a meta-analytic review. *Perspect Psychol Sci.* 2016;11(4):512–534. <https://doi.org/10.1177/1745691616635612>.
- Moore ME, Ross BM. Context effects in running memory. *Psychol Rep.* 1963;12:451–465.
- Myers NE, Chekroud SR, Stokes MG, Nobre AC. Benefits of flexible prioritization in working memory can arise without costs. 2018;44(3):398–411.
- Norris DG, Hall J, Gathercole SE. Can short-term memory be trained? *Mem Cogn.* 2019;47(5):1012–1023. <https://doi.org/10.3758/s13421-019-00901-z>.
- Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc Natl Acad Sci U S A.* 2001;98(2):676–682. <https://doi.org/10.1073/pnas.98.2.676>.
- Ritakallio L, Fellman D, Jylkkä J, Waris O, Lönnroth N, Nervaander R, et al. The pursuit of effective working memory training: a pre-registered randomised controlled trial with a novel varied training protocol. *J Cogn Enhanc.* 2021:1–16. <https://doi.org/10.1007/s41465-021-00235-2>.
- Rodriguez-Jimenez R, Avila C, Garcia-Navarro C, Bagney A, Aragon AM D, Ventura-Campos N, et al. Differential dorsolateral prefrontal cortex activation during a verbal n-back task according to sensory modality. *Behav Brain Res.* 2009;205(1):299–302. <https://doi.org/10.1016/j.bbr.2009.08.022>.
- Shipstead Z, Redick TS, Engle RW. Does working memory training generalize? *Psychol Belg.* 2010;50(3–4):245. <https://doi.org/10.5334/pb-50-3-4-245>.
- Shipstead Z, Redick TS, Engle RW. Is working memory training effective? *Psychol Bull.* 2012;138(4):628–654. <https://doi.org/10.1037/a0027473>.
- Shulman GL, Corbetta M, Buckner RL, Fiez JA, Miezin FM, Raichle ME, Petersen SE. Common blood flow changes across visual tasks: I. Increases in subcortical structures and cerebellum but not in nonvisual cortex. *J Cogn Neurosci.* 1997;9(5):624–647. <https://doi.org/10.1162/jocn.1997.9.5.624>.
- Swanson HL. Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Mem Cogn.* 2004;32(4):648–661. <https://doi.org/10.3758/BF03195856>.
- Todd JJ, Fougny D, Marois R. Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentive blindness. *Psychol Sci.* 2005;16(12):965–972. <https://doi.org/10.1111/j.1467-9280.2005.01645.x>.
- Unsworth N, Heitz RP, Schrock J. C., Engle R. W. An automated version of the operation span task. *Behavior research methods.* 2005;37(3):498–505.
- Velan H, Frost R. Cambridge University versus Hebrew University: the impact of letter transposition on reading English and Hebrew. *Psychon Bull Rev.* 2007;14(5):913–918. <https://doi.org/10.3758/BF03194121>.
- Wechsler D, Coalson D, Raiford S IV. WAIS-III Wechsler adult intelligence scale San Antonio. TX: Psychological Corporation; 1997
- Zatorre, Belin. Spectral and temporal processing in human auditory cortex. *Cereb Cortex.* 2001;11(10):946–953. <https://doi.org/10.1093/cercor/11.10.946>.
- Zatorre RJ, Halpern AR, Bouffard M. Mental reversal of imagined melodies: a role for the posterior parietal cortex. *J Cogn Neurosci.* 2010;22(4):775–789. <https://doi.org/10.1162/jocn.2009.21239>.
- Zatorre R, Delhommeau K, Zarate J. Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Front Psychol.* 2012;3(DEC):1–17. <https://doi.org/10.3389/fpsyg.2012.00544>.