Memory & Cognition
https://doi.org/10.3758/s13421-024-01530-x

Is *Hey Jude* in the right key? Cognitive components of absolute pitch memory

Stephen C. Van Hedger[1,2] · **Andrea R. Halpern**[3] · **David J. Vollweiler**[4] · **Evan E. Smith**[1] · **Peter Q. Pfordresher**[4]

Accepted: 19 January 2024

Abstract

Most individuals, regardless of formal musical training, have long-term absolute pitch memory (APM) for familiar musical recordings, though with varying levels of accuracy. The present study followed up on recent evidence suggesting an association between singing accuracy and APM (Halpern & Pfordresher, 2022, *Attention, Perception, & Psychophysics*, *84*(1), 260–269), as well as tonal short-term memory (STM) and APM (Van Hedger et al., 2018, *Quarterly Journal of Experimental Psychology*, *71*(4), 879–891). Participants from three research sites ($n = 108$) completed a battery of tasks including APM, tonal STM, singing accuracy, and self-reported auditory imagery. Both tonal STM and singing accuracy predicted APM, replicating prior results. Tonal STM also predicted singing accuracy, music training, and auditory imagery. Further tests suggested that the association between APM and singing accuracy was fully mediated by tonal STM. This pattern comports well with models of vocal pitch matching that include STM for pitch as a mechanism for sensorimotor translation.

Keywords Pitch · Long-term memory · Short-term memory · Singing · Imagery

Introduction

The typical laboratory study of memory accuracy often involves intentional encoding of new and arbitrary information, like lists of words or novel faces. However, in everyday life, most information is encoded incidentally and in a real-world context. Accuracy of these memories is harder to assess. For instance, in autobiographical memory, qualities like vividness and emotionality can be assessed, but validating the reported experience is often not possible.

One type of memory that is incidental, meaningful, and can be assessed is absolute pitch memory (APM). This is the ability to produce or recognize the starting pitches of a familiar piece of music that has only ever been heard in one key, without any musical context or explicit knowledge of the note names (i.e., not having looked at the sheet music). APM is distinct from being able to name a sounded note (e.g., "B flat") or produce a note given its name, called absolute pitch (AP). APM is more accurate and widespread than many would predict. For instance, in a large-scale replication of a prior study (Levitin, 1994), Frieler et al. (2013) found that a quarter of their sample of 277 people, over six labs, were able to sing the precise opening note for at least one of two self-selected pop recordings. That study and others (e.g., Schellenberg & Trehub, 2003; Van Hedger et al., 2018, 2023) show a distribution of accuracy centered on the correct note, whether assessed by vocalizing the opening note, finding the note on a keyboard, or recognizing whether a track has been pitch-shifted by a small amount.

Our focus here was on that variability among people. Van Hedger et al. (2018), using a recognition task, found mean accuracy of .61 (.50 = chance) but with a standard deviation of .10 for familiar recordings in Experiment 1. Halpern and Pfordresher (2022) asked participants to find the opening pitch of 10 familiar songs on a keyboard. Whereas 13 of the 46 participants averaged no more than 1 semitone error, performance ranged from an

✉ Stephen C. Van Hedger
svanhedg@uwo.ca

[1] Department of Psychology, Huron University College at Western, London, ON, Canada

[2] Department of Psychology and Western Institute for Neuroscience, Western University, London, ON, Canada

[3] Department of Psychology, Bucknell University, Lewisburg, PA, USA

[4] Department of Psychology, University at Buffalo, The State University of New York, Buffalo, NY, USA

average of 7 semitones flat to 12 semitones sharp. Frieler et al. (2013) found a mean error of vocalized opening note of about 2.6 semitones flat, but with a standard deviation of 8 semitones.

This variability invites the question of individual differences in component skills that might account for the wide range of performance on this task. One obvious candidate is musical training, given that training hones auditory discrimination and memory skills (e.g., Barrett et al., 2013). However, years of musical training failed to predict APM in any of the studies referenced above.

In contrast, several studies have found interesting correlational relationships with *cognitive and perceptual* variables. Jakubowski and Müllensiefen (2013) showed that a test of relative pitch memory (is Tune 2 the same as Tune 1) correlated with vocal production APM, implying a role of short-term pitch memory in the ability to encode long-term APM. In a direct test of this idea, Van Hedger et al. (2018) administered a tonal short-term memory (STM) task[1] in which a participant hears a sine wave target, then a masking noise, then adjusts a starting tone to reproduce the target by moving arrows on a screen that change the pitch by small amounts each time. This task (but not verbal working memory, measured through an auditory *n*-back) predicted recognition APM performance, providing evidence that accuracy in pitch memory representation generalizes over short and long-term retention, as well as intentional and incidental encoding contexts.

Another individual difference variable related to APM is more *sensorimotor*: vocal pitch matching. Like APM, many people can accurately mimic a target note or short pattern, but with large variability across individuals. Vocal pitch matching, a simple task at first glance, requires that several processes be executed quickly and precisely: perceiving then generating an auditory image of the target pitch in auditory memory, forming a plan to match the target using muscles not normally visible, execution of the plan, and evaluation. Focusing on image generation, Pfordresher and Halpern (2013) found that at least perceived fidelity in imagining sounds—namely, auditory imagery vividness (measured by the Bucknell Auditory Imagery Scale [BAIS]; Halpern, 2015), predicted singing accuracy. Other studies have linked auditory imagery task performance to pitch matching. For instance, Greenspon and Pfordresher (2019) found that accuracy in pitch matching correlated with memory for imagined musical pitches. Relevant to the current question, Halpern and Pfordresher (2022) found that singing accuracy predicted keyboard production APM (note, participants are prohibited from humming during the task). However, auditory imagery vividness did not predict keyboard APM, and auditory imagery clarity (measured by the Clarity of Auditory Imagery Scale; Willander & Baraldi, 2010) likewise did not predict recognition APM in the Van Hedger et al. (2018) study.

This set of findings thus implies a model whereby this variety of incidental nonsemantic memory might be facilitated by having well-functioning auditory imagery and STM components. However, prior studies, including from our labs, used only a subset of predictors and used different tests of APM. Therefore, in the current study, we systematically explored the contributions of auditory imagery self-report, singing accuracy, and accuracy of tonal STM to APM. We used the recognition version of APM both to allow us to generalize findings of Halpern and Pfordresher (2022) in relating singing accuracy to a task not involving any kind of production and also to provide a replication of the tonal STM–APM link found in Van Hedger et al. (2018). We included the BAIS partly to validate the relationship of that to singing accuracy and to test if the lack of direct relationship of self-reported imagery to APM replicates.

Data were collected at three sites, to facilitate recruitment of a sample sufficient to model how the predictors might interact in explaining APM variance. We hypothesized that auditory imagery would predict APM but be mediated through singing accuracy (a sensorimotor component), but that accuracy of tonal STM (a cognitive component) would predict APM directly.

Method

Participants

A total of 108 participants (Age: M = 19.60 years, SD = 4.19, range: 18–57[2] years; 70 females, 38 males) across three sites (Bucknell University: n = 35; Huron University College: n = 41; University at Buffalo: n = 32) were included in the primary analyses. The sample size was determined a priori in terms of minimum participant contributions (n = 30) for each site, as this would provide adequate power ($1 - \beta = .80$) for detecting medium-sized (r = .30) correlations, which were expected given prior work (Halpern & Pfordresher, 2022; Van Hedger et al., 2018). The final sample size, which exceeded these minimum recruitment

---

[1] The TSTM task in Van Hedger et al. (2018) was called the implicit note memory (INM) task. However, it was treated as a measure of tonal short-term memory precision.

[2] The oldest participant (57 years old) was an outlier in terms of age and was substantially older than the next oldest participant (34 years old). However, we opted to include the oldest participant as they were not outliers—defined by the Interquartile Range (IQR) Method of 1.5*IQR—on any of the measures.

goals, had .90 power to detect medium-sized correlations. Participants with AP were excluded from the study. Each site received institutional ethics approval.

Materials

The tonal short-term memory (TSTM) task was modeled on Van Hedger et al. (2018) and required participants to adjust a starting note to match an initially presented target note. On each trial participants listened to a single *target note* (either F#4, G4, G#4, or A4) followed by 1,000 ms of masking noise. Participants then adjusted a *starting note* to match the target note by clicking on up and down arrows displayed on the screen, which increased or decreased the starting tone by one-third of a semitone, respectively. There were eight possible starting notes: four below the range of the target notes (D4, D#4, E4, and F4) and four above the range of the target notes (A#4, B4, C#5, and C#5). The target note could not be replayed during the trial. All tones were 250 ms in duration. Once participants were satisfied with their response, they pressed a designated key to continue to the next trial. Participants were given two initial practice trials, which were not scored, to become familiar with the task. In the main assessment, there were 64 trials with every combination of target notes and starting notes included twice.

TSTM performance was operationalized as the mean absolute deviation from the target note, represented in terms of arrow clicks (e.g., a mean score of 5 would reflect that a participant was, on average, five clicks or 1.67 semitones away from the actual target note). Individual trials that were more than 3 standard deviations away from a participant's mean were discarded, as outlined in Van Hedger et al. (2018). The TSTM task was coded in jsPsych 7 (de Leeuw, 2015).

The absolute pitch memory task (APM) task, also modeled on Van Hedger et al. (2018), required participants to judge whether short excerpts from musical recordings (5 s in duration with a 500-ms fade-in and fade-out) had been shifted in pitch. The 28 excerpts were confirmed to be highly recognizable through previous pilot testing and for which any professional covers were in the same key as the original (see the Appendix for a list of recordings). There were 14 correct (i.e., unshifted) and 14 incorrect (i.e., shifted) trials, presented in a random order. Of the 14 incorrect trials, seven were shifted by +1 semitone and seven were shifted by −1 semitone. To ensure that no cues were available from the editing procedure, correct stimuli were shifted up in pitch by 0.5 semitones and then down by 0.5 semitones. Incorrect +1 semitone stimuli were shifted up in pitch by 0.5 semitones twice. Incorrect −1 semitone stimuli were shifted down in pitch by 0.5 semitones twice. Pitch shifting was done in Audacity using a "high-quality stretching" option to preserve the length of each recording. The assignment of an excerpt to be either correct, incorrect by +1 semitone, or incorrect by −1 semitone was randomized across participants.

For each excerpt, participants indicated via button press whether the song sounded correct (yes/no). If participants selected "no," they indicated whether the recording sounded "too high" or "too low" in pitch. Participants were also asked to indicate how familiar the recording was to them on a Likert-type scale via button press (*Not at all*, *A little*, *Somewhat*, *Quite a bit*, *Extremely*). Trials were discarded if participants reported no prior familiarity with the recording because the task assumes familiarity with the recording (cf. Schellenberg & Trehub, 2003). In the present experiment, this procedure removed only 7.5% of trials. APM performance was operationalized in terms of proportion correct for the initial question of whether the recording sounded correct. The follow-up question asking participants to indicate whether they thought a recording sounded "too high" or "too low" is novel to the present study, and as such was treated as an exploratory measure of APM (reported in Supplemental Material).

The Bucknell Auditory Imagery Scale (Halpern, 2015) assessed participants' self-reported auditory imagery on two dimensions: vividness (14 items), and control (11 items[3]). For auditory imagery vividness, participants rated how vividly they could think of an auditory image in their head for several scenarios (e.g., "For the next item, consider ordering something over the phone. [How vivid is] the voice of an elderly clerk assisting you?"). Ratings were made on a 7-point Likert-type scale (1 = *No image present at all* to 7 = *As vivid as the actual sound*). For auditory imagery control, participants were given pairs of scenarios that required them to change a sound in their mind's ear and were subsequently asked how easily they were able to do so (e.g., "Consider attending a choir rehearsal. (a) The sound of an all-children's choir singing the first verse. (b) The sound of an all-adults' choir now sings the second verse of the song"). Like vividness, participants made their ease-of-change ratings on a 7-point Likert-type scale (1 = *No image present at all* to 7 = *Extremely easy to change the image*). We measured individual differences using mean scores for each participant on each subscale. In the present study, the vividness and control scales were significantly correlated with one another ($r = .63$), and both displayed good internal consistency (vividness: $\alpha = .79$; control: $\alpha = .76$). The Bucknell Auditory Imagery Scale was programmed in Qualtrics (Provo, UT).

The Seattle Singing Accuracy Protocol assessed accuracy of vocal pitch matching. After completing a warmup exercise, participants completed 26 scored trials. The first

---

[3] The final three items of the Clarity Subscale of the BAIS were unintentionally not administered to participants.

10 trials required participants to reproduce a single target note based on a vocal timbre. The next 10 trials required participants to reproduce a single target note based on a piano timbre. The final six trials required participants to reproduce four-note melodies based on a vocal timbre. Pitch ranges for the target sounds were determined by participant's self-reported vocal gender and range of pitches used during warmup. Pitch accuracy in reproduction was based on whether the mean f_0 value of imitated pitches (excluding outliers and starting/ending portions) was within ±50 cents (1/2 semitone) of the corresponding target pitch (for details, see Pfordresher & Demorest, 2020). The Seattle Singing Accuracy Protocol was programmed in MATLAB (The MathWorks, Natick, MA).

The specific testing hardware varied by site. At Bucknell, participants were tested on a Mac Pro, played via Advent AVS 570 speakers, and were vocally recorded with the internal microphone on the computer. At Huron University College, participants were tested on a Lenovo ThinkPad T480 laptop. Participants listened to the sounds through Sony MDR-7506 stereo professional headphones and were recorded on an Audio-Technica AT2020 stereo microphone. Both the headphones and microphone were connected to a Steinberg UR12 USB audio interface. At the University at Buffalo, participants were tested on a Dell Precision PC, listened to the sounds through Sennheiser HD 280 Pro headphones, and were vocally recorded with a Shure SM58, connected through a Lexicon Omega I/O USB box.

Procedure

Upon providing informed consent, participants first completed the TSTM and APM tasks. The ordering of the TSTM and APM was varied across participants. After completion of both the TSTM and APM, participants were redirected to Qualtrics to complete the Bucknell Auditory Imagery Scale (vividness and control, in this order). There were four demographic questions (age, sex, years of education, and occupation) that preceded the scale.

Following the Bucknell Auditory Imagery Scale, participants completed the Seattle Singing Accuracy Protocol. The warmup exercise required participants to sing a familiar song, chosen from a prespecified list, on the syllable "doo." Following the 26 scored trials, participants re-sung the familiar song they selected in warmup with the song's lyrics (provided on the computer screen). The script additionally included a pitch discrimination task based on Loui et al. (2008) and a short hearing, music, and language experience questionnaire. Years of musical lessons (summed across all reported instruments, including voice) was used as a measure of musical training. Following the Seattle Singing Accuracy Protocol, participants were debriefed and provided with compensation (monetary or course credit) depending on the approved ethics protocols from each institution.

Data analysis

Data were analyzed in R (Version 4.1.2). Performance on the APM task was assessed via a one-sample t test against the chance estimate of .50. This corresponded to trials in which participants reported at least some familiarity with the recording. Bivariate associations among variables were assessed via Pearson product-moment correlations.

We used hierarchical regression to assess how the measured variables related to APM performance. We opted to use hierarchical regression given that the measured variables had differing predicted association strengths with APM, and this approach allowed us to enter measured variables in an informed manner (and hierarchical regression has been used in conceptually similar work; e.g., Colley et al., 2018; Greenspon & Pfordresher, 2019). The hierarchical regression models used generalized linear models, as the dependent variable was the proportion of correct trials. Each model used a quasibinomial link and was weighted based on the total of included APM trials (i.e., trials in which participants reported at least some familiarity with the recording). The first step (relative to a null model) included data collection site to control for mean performance differences across site. The second step added musical training as another control, (cf. Greenspon & Pfordresher, 2019), as musical training was not expected to directly relate to APM but was expected to relate to TSTM (cf. Van Hedger et al., 2018). The third and fourth steps added TSTM and singing accuracy, respectively, as these have been previously associated with APM (Halpern & Pfordresher, 2022; Van Hedger et al., 2018). TSTM was added before singing accuracy because short-term and working memory has been implicated more broadly in differentiating performance among AP possessors (Deutsch & Dooley, 2013; Van Hedger & Nusbaum, 2018) and explaining individual variability in explicit category learning of AP (Van Hedger et al., 2015), in addition to being associated with individual differences in APM (Van Hedger et al., 2018). The contribution of singing accuracy to APM, in contrast, is less firmly established and could be considered more exploratory (Halpern & Pfordresher, 2022). The fifth and sixth steps added the reported vividness and control of auditory imagery in this order (cf. Greenspon & Pfordresher, 2019). Vividness was added before control because prior work has found singing accuracy to relate to reported vividness of auditory imagery (Halpern & Pfordresher, 2022). These nested models were compared through chi-squared tests using the "anova" function in R. Given that the models were generalized linear models, goodness-of-fit was assessed through calculating the change in deviance from each step (with a significant reduction in deviance indicating a better fit).

Mediation analyses used the "mediation" package in R (Tingley et al., 2014). The average causal mediation effect and direct effect were estimated through a bootstrapping procedure (5,000 simulations). The significance of each effect

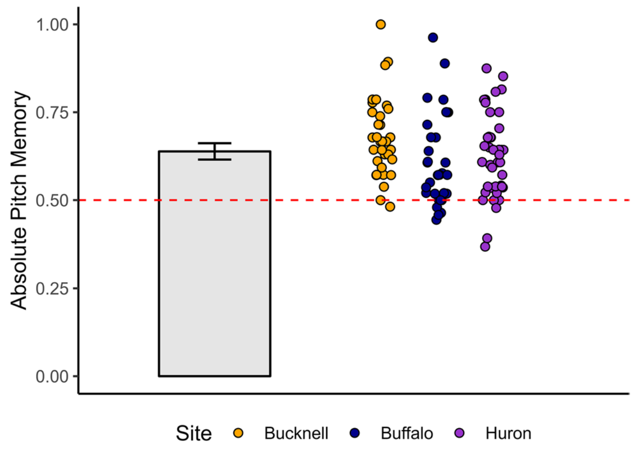

Fig. 1 Performance on the absolute pitch memory (APM) task. *Note*. The bar on the left represents aggregate performance across all participants. The error bar represents the 95% confidence interval of the mean. Individual data points are grouped by recruitment site and plotted on the right side of the figure. The dashed line represents chance performance. (Color figure online)

was determined by assessing whether the 95% confidence interval from the bootstrapping procedure included zero.

Results

Testing APM performance against chance

Participants selected the correct response 63.9% of the time (*SD* = 12.2%), which was significantly above the chance estimate, $t(107) = 11.81$, $p < .001$, $d = 1.14$. Figure 1 displays a visualization of performance on the APM measure across site.

Bivariate correlations of variables

Correlations among the measured variables are reported in Table 1. APM was significantly correlated with TSTM, with higher APM scores associated with smaller deviations in reproducing a target tone (i.e., better TSTM). APM performance was also significantly correlated with singing accuracy, with higher APM scores associated with more accurate singing. APM performance was also significantly associated with the reported control of auditory imagery, with higher APM scores associated with greater amounts of self-reported control of auditory imagery. These three significant correlations are plotted in Fig. 2. APM performance was not significantly associated with musical training or the reported vividness of auditory imagery.

There were additionally several observed interrelationships among the non-APM measures. TSTM was strongly associated with singing accuracy, with better TSTM associated with better singing accuracy. Better TSTM was additionally associated with greater amounts of musical training, as well as with higher scores on both auditory imagery vividness and control. Singing accuracy was significantly associated with the reported vividness of auditory imagery and musical training, but not with the reported control of auditory imagery.

Hierarchical regression

Summary results of the hierarchical regression analyses are provided in Table 2. The first two steps of adding site and musical training as control variables did not significantly improve model fit relative to the null model. The third step of adding TSTM resulted in a significantly improved model fit, with TSTM acting as the sole significant predictor of

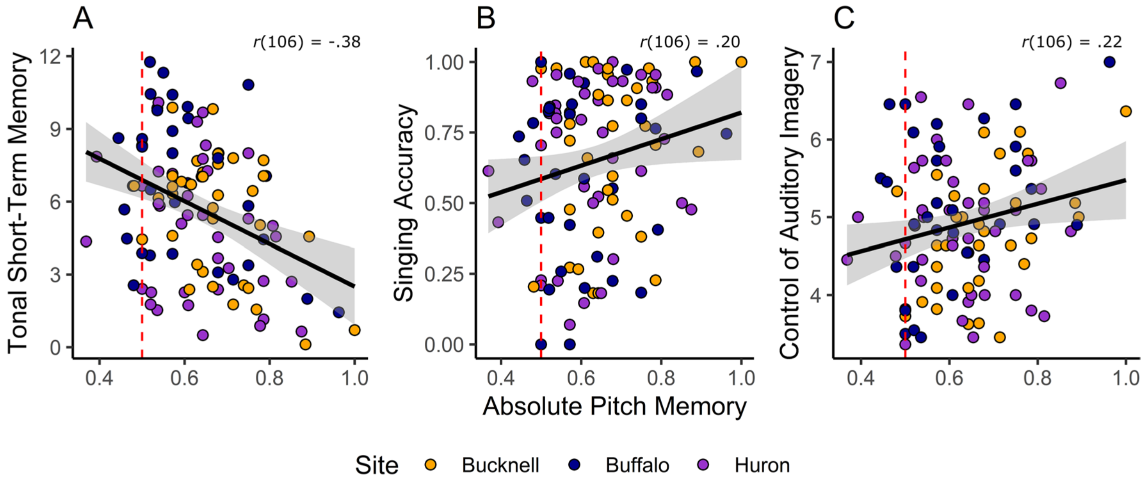

Fig. 2 Significant correlations between absolute pitch memory and tonal short-term memory (**A**), singing accuracy (**B**), and the control of auditory imagery (**C**). *Note*. Error ribbons represent 95% confidence intervals. Individual data points are colored based on recruitment site. The dashed red line represents chance performance on the APM task. (Color figure online)

Table 1 Correlation matrix for measured variables

Measure	APM	TSTM	SSAP	BAIS-C	BAIS-V
APM	–				
TSTM	−.38***	–			
SSAP	.20*	−.57***	–		
BAIS-C	.22*	−.31***	.18	–	
BAIS-V	.13	−.31***	.19*	.63***	–
Music training	.17	−.36***	.35***	−.08	.05

APM = absolute pitch memory; TSTM = tonal short-term memory; SSAP = Seattle Singing Accuracy Protocol; BAIS-C = Bucknell Auditory Imagery Scale (Control Subscale); BAIS-V = Bucknell Auditory Imagery Scale (Vividness Subscale). ***p < .001, **p < .01, * p < .05

APM. The subsequent steps of adding singing accuracy (fourth step), reported vividness of auditory imagery (fifth step), and reported control of auditory imagery (sixth step) did not significantly improve the model fit. In the final model incorporating all variables, TSTM was significant, whereas singing accuracy, music training, reported vividness of auditory imagery, and reported control of auditory imagery were all nonsignificant (*p*s > .185).

Mediation analyses

The predicted mediation path—that singing accuracy would mediate the relationship between self-reported auditory imagery and APM—was not supported for either the reported vividness or control of auditory imagery. These results are perhaps not surprising, as the preconditions for mediation were not fully satisfied for either the reported vividness of auditory imagery (nonsignificant association to APM) or the reported control of auditory imagery (nonsignificant association to singing accuracy). However, the regression analyses supported another possible mediation path in which TSTM mediates the relationship between singing accuracy and APM, as well as the relationship between the reported control of auditory imagery and APM. Although not initially predicted, models of singing accuracy suggest that TSTM and working memory play a role within the singing system (e.g., the MMIA model; Pfordresher et al., 2015). If TSTM is necessary for singing accuracy, a reasonable hypothesis is that TSTM may mediate any relationship between singing accuracy and APM. Additionally, the control of auditory imagery has face validity in being associated with short-term and working memory, as both involve not only maintenance of an image but also dynamic mental comparison (for instance, both working memory and the control of auditory imagery predicted the ability to predict the onset of the beat in expressively timed music; Colley et al., 2018).

The preconditions for these mediations were satisfied, as APM was significantly associated with TSTM, singing accuracy, and the reported control of auditory imagery; likewise, TSTM was significantly correlated with singing accuracy and the reported control of auditory imagery. For singing accuracy, the average causal mediation effect was 0.08, with the 95% confidence interval not including zero (0.03, 0.13). The direct effect (i.e., the effect of singing accuracy on APM when accounting for TSTM) was −0.01, with the 95% confidence interval including zero (−0.08, 0.07), suggesting that TSTM robustly mediated the relationship between singing accuracy and APM. For the reported control of auditory imagery, the average causal mediation effect was 0.02, with the 95% confidence interval not including zero (0.01, 0.03). The direct effect (i.e., the effect of reported auditory imagery control on APM when accounting for TSTM) was 0.02, with the 95% confidence interval including zero (−0.01, 0.05), suggesting that TSTM fully mediated the relationship between the reported control of auditory imagery and APM. The mediation paths for both singing accuracy and the reported control of auditory imagery are plotted in Fig. 3.

Table 2 Summary of hierarchical regression model comparisons

Step/Predictor	Step 1 β	Step 2 β	Step 3 β	Step 4 β	Step 5 β	Step 6 β
Site: Buffalo	−.16*	−.13	−.11	−.11	−.12	−.14
Site: Huron	−.12	−.11	−.14	−.14	−.14	−.14
Music Training		.04	.00	.00	.00	.00
TSTM			−.13***	−.13***	−.12**	−.12**
SSAP				−.01	−.01	−.01
BAIS-V					.02	−.01
BAIS-C						.06
Pseudo R^2	.04	.05	.17	.17	.17	.18
Model deviance	189.47	187.38	164.42***	164.37	163.98	161.11

TSTM = tonal short-term memory; SSAP = Seattle Singing Accuracy Protocol; BAIS = Bucknell Auditory Imagery Scale. Site was dummy coded, with Bucknell serving as the reference category. Lower TSTM scores represent better performance, whereas higher scores on the SSAP, BAIS-C, and BAIS-V represent better performance on these measures. The only step that resulted in a significantly better goodness-of-fit (lower model deviance) was Step 1, in which TSTM was added. Betas represent standardized coefficients. ***p < .001 **p < .01 *p < .05

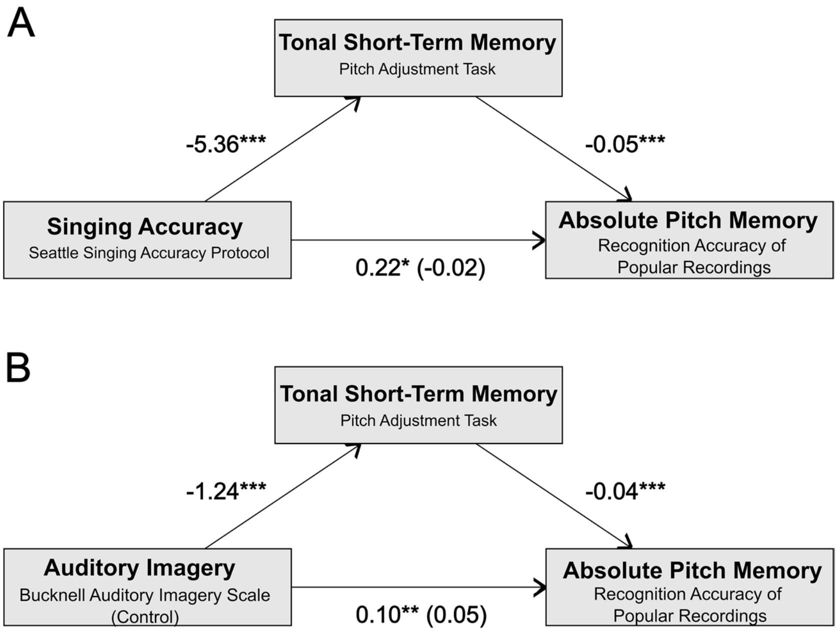

Fig. 3 Mediation paths demonstrating that tonal short-term memory fully mediated the relationship between absolute pitch memory and singing accuracy (**A**), as well as absolute pitch memory and control of auditory imagery (**B**). *Note*. Values represent unstandardized regression coefficients. The values in parentheses represent the regression coefficient when tonal short-term memory is included in the model. ***$p < .001$ **$p < .01$ *$p < .05$.

Discussion

We report the first evidence here linking absolute pitch memory (APM) for familiar songs to tonal short-term memory (TSTM), singing accuracy, and reported ability to control auditory imagery. As such, this research provides evidence that APM may arise from the functioning of a sensorimotor network. Mediation and hierarchical regression analyses further suggest that accuracy of TSTM acts as the foundation for this network, through which individual differences in singing accuracy and controllability of auditory imagery are associated with APM. Consistent with prior work (e.g., Van Hedger et al., 2018), musical training was not directly associated with APM, despite being strongly correlated with TSTM and singing accuracy. These results shed light on the mechanisms that underlie memory and imagery and suggest possibilities for remediation techniques.

The nature of long-term memories for musical stimuli has intrigued researchers for decades (e.g., Dowling & Bartlett, 1981; Halpern, 1989). Recent research by Van Hedger and colleagues (2018) found that APM is associated with the accuracy of TSTM, whereas Halpern and Pfordresher (2022) illustrated an association between keyboard APM and singing accuracy. The current results add to this by showing that this effect is mediated by TSTM, thus suggesting that the ability to remember pitches accurately and precisely over the short term facilitates encoding of precise absolute pitch over the long term. This interpretation is consistent with recent perspectives in vision research that short- and long-term memory precision neurally overlap (Ester et al., 2013), may have the same fidelity constraints (Brady et al., 2013; though see Biderman et al., 2019), and, most pertinently for the present work, are significantly correlated constructs within individuals (Xie et al., 2020).

This research helps clarify the role of pitch memory in singing accuracy (i.e., vocal pitch matching). The present study revealed significant associations between singing accuracy and two measures of recognition memory, one long and one short-term. This provides strong evidence for previous claims that singing may draw on cognitive resources that serve general auditory cognition and are not specific to motor control (e.g., Pfordresher et al., 2015). The present data conceptually replicate earlier associations between singing accuracy with an APM task based on production (Halpern & Pfordresher, 2022). Additionally, the present data conceptually replicate an earlier reported association between singing accuracy and short-term memory for pitch (Greenspon & Pfordresher, 2019); however, the present study found a considerably larger association between tonal STM with singing accuracy than this earlier work ($r^2 = .32$ in the current study versus $r^2 = .06$ in Greenspon & Pfordresher, 2019). It is potentially significant that the current STM task relies on precise reproductions of single pitches, drawing on pitch memory for single tones, whereas the previous measure (from Williamson & Stewart, 2010) was a span task, which requires processing the serial position of discrete pitches. The potential importance of precise pitch associations relates also to previous correlations between singing accuracy and reproducing pitch using a slider (Demorest & Clements, 2007; Hutchins & Peretz, 2012).

The present data also shed light on the role of auditory imagery in recognition and production. As in Pfordresher and Halpern (2013), singing accuracy correlated with self-reported vividness of imagery, but not with self-reported control. By contrast, APM correlated with reported control but not with reported vividness of auditory imagery. Mediation analysis further suggested that the correlation of APM with reported auditory imagery, like singing accuracy, was fully mediated by TSTM. These results verify that the two subtests of the Bucknell Auditory Imagery Scale, though strongly correlated with each other, reflect distinct processes that serve different purposes within auditory cognition. Although all the items in the Bucknell Auditory Imagery Scale draw on long-term memory, the control subtest also draws heavily on working memory by having users reflect on the distinction between an initially formed image and a later transformed image. Correlations between vividness and singing accuracy, found here and elsewhere, may reflect the critical importance of initial imagery formation in motor planning (see also Greenspon et al., 2017).

In exploring the relationship between vocal pitch matching and APM captured by recognition, the present research selected measures based on significant associations revealed in recent studies. Of course, one could explore other associations given the complexities underlying these behaviors, including associations between the capacity of auditory short-term memory for pitch (cf. Williamson & Stewart, 2010), and associations with auditory pitch perception and objective measures of pitch imagery (e.g., Greenspon & Pfordresher, 2019). Additionally, here we focused on one measure of singing accuracy—percentage of correctly matched pitches—which is how accuracy is often defined. Nevertheless, during the analysis phase, we also examined bivariate correlations between memory measures and the more nuanced measures of mean absolute difference between each sung and each target pitch, as well as the precision or consistency of pitch production (following Berkowska & Dalla Bella, 2013; Pfordresher et al., 2010). These measures yielded similar patterns of correlation with the two memory measures, but with smaller r values. One exception was that pitch precision did not correlate significantly with APM ($r = -.11$). We are confident that similar conclusions would emerge from different measures to assess vocal pitch matching and pitch memory.

In conclusion, the ability to recognize the correct absolute pitch content of a familiar melody is positively associated with both one's ability to reproduce pitches in that melody via singing, the ability to form and manipulate an auditory image of that melody, and how precisely one can sustain the content of individual pitches in TSTM. Of all these associations, the link between TSTM and APM is foundational. TSTM mediates associations between APM and individual differences in both singing accuracy and imagery control. The central role of TSTM in the present study mirrors the critical importance of short-term and working memory in other domains, such as category learning (e.g., Hambrick et al., 2004; Lewandowsky, 2011). Although evidence is weak for musical training yielding far transfer effects to nonmusical domains (e.g., Schellenberg & Lima, 2024), our observed association between TSTM and singing accuracy raises the possibility that remediation of poor-pitch singing may benefit from training of short-term or working memory for pitch. However, to our knowledge, this idea has not been empirically tested and could represent a promising avenue for future research. Even more foundationally, this work suggests that memory accuracy may be a reliable individual difference that generalizes over short- and long-term memory, as well as over radically different encoding contexts, and complexity of the materials.

Appendix

Title	Artist
Single Ladies	Beyoncé
Umbrella	Rihanna
Shake It off	Taylor Swift
Toxic	Britney Spears
Rolling in the Deep	Adele
Firework	Katy Perry
Blinding Lights	The Weeknd
Hey Ya!	Outkast
Hips Don't Lie	Shakira
Bringing Sexy Back	Justin Timberlake
Call Me Maybe	Carly Rae Jepsen
Uptown Funk	Bruno Mars
Poker Face	Lady Gaga
Starships	Nicki Minaj
Royals	Lorde
Party in the U.S.A.	Miley Cyrus
bad guy	Billie Eilish
Get Lucky	Daft Punk
Happy	Pharrell Williams
Despacito	Luis Fonsi
Gangnam Style	PSY
Take On Me	a-ha
Sweet Child O' Mine	Guns N' Roses
Imagine	John Lennon
We Are the Champions	Queen
Smells Like Teen Spirit	Nirvana
Somebody That I Used to Know	Gotye
American Pie	Don McLean

28 excerpts of popular songs that were used in the APM task.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-024-01530-x.

Acknowledgements We thank Abdullah Alam, Lihui De, Mitchell Dugan, Emma Gandolfi, Elizabeth Goldstein, Michaela Lester, and Richard Nicosia for help with data collection.

Data Availability All data and analysis scripts are available through Open Science Framework (https://doi.org/10.17605/OSF.IO/QP5RW).

Declarations

Conflicts of interest We have no conflicts of interest to declare.

References

Barrett, K. C., Ashley, R., Strait, D. L., & Kraus, N. (2013). Art and science: How musical training shapes the brain. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00713

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The Sung Performance Battery (SPB). *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00714

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working memory has better fidelity than long-term memory: The fidelity constraint is not a general property of memory after all. *Psychological Science, 30*(2), 223–237. https://doi.org/10.1177/0956797618813538

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*(6), 981–990. https://doi.org/10.1177/0956797612465439

Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music. *Quarterly Journal of Experimental Psychology, 71*(8), 1781–1796. https://doi.org/10.1080/17470218.2017.1366531

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Demorest, S. M., & Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *Journal of Research in Music Education, 55*(3), 190–203. https://doi.org/10.1177/002242940705500302

Deutsch, D., & Dooley, K. (2013). Absolute pitch is associated with a large auditory digit span: A clue to its genesis. *The Journal of the Acoustical Society of America, 133*(4), 1859–1861. https://doi.org/10.1121/1.4792217

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology: A Journal of Research in Music Cognition*, *1*(1), 30–49. https://doi.org/10.1037/h0094275

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience, 25*(5), 754–761. https://doi.org/10.1162/jocn_a_00357

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Müllensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae, 17*(3), 334–349. https://doi.org/10.1177/1029864913493802

Greenspon, E. B., & Pfordresher, P. Q. (2019). Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation. *Attention, Perception, & Psychophysics, 81*(7), 2473–2481. https://doi.org/10.3758/s13414-019-01799-0

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception, 34*(5), 585–604. https://doi.org/10.1525/mp.2017.34.5.585

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17*(5), 572–581. https://doi.org/10.3758/BF03197080

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*(1), 37–47. https://doi.org/10.1037/pmu0000081

Halpern, A. R., & Pfordresher, P. Q. (2022). What do less accurate singers remember? Pitch-matching ability and long-term memory for music. *Attention, Perception, & Psychophysics, 84*(1), 260–269. https://doi.org/10.3758/s13414-021-02391-1

Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2004). The role of working memory in higher-level cognition: Domain-specific versus domain-general perspectives. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence* (pp. 104–121). Cambridge University Press. https://doi.org/10.1017/CBO9780511607073.007

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General, 141*(1), 76–97. https://doi.org/10.1037/a0025064

Jakubowski, K., & Müllensiefen, D. (2013). The influence of music-elicited emotions and relative pitch on absolute pitch memory for familiar melodies. *Quarterly Journal of Experimental Psychology, 66*(7), 1259–1267. https://doi.org/10.1080/17470218.2013.803136

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56*(4), 414–423. https://doi.org/10.3758/BF03206733

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 720–738. https://doi.org/10.1037/a0022639

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology, 18*(8), R331–R332. https://doi.org/10.1016/j.cub.2008.02.045

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America, 128*(4), 2182–2190. https://doi.org/10.1121/1.3478782

Pfordresher, P. Q., & Demorest, S. M. (2020). Construction and validation of the Seattle Singing Accuracy Protocol (SSAP). In F. A. Russo, B. Ilari, & A. J. Cohen (Eds.), *The Routledge companion to interdisciplinary studies in singing* (pp. 322–333). Routledge. https://doi.org/10.4324/9781315163734-24

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review, 20*(4), 747–753. https://doi.org/10.3758/s13423-013-0401-8

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing. *Music Perception, 32*(3), 242–253. https://doi.org/10.1525/mp.2015.32.3.242

Schellenberg, E. G., & Lima, C.F. (2024). Musical training and non-musical abilities. *Annual Review of Psychology*, *75*(1). https://doi.org/10.1146/annurev-psych-032323-051354

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14*(3), 262–266. https://doi.org/10.1111/1467-9280.03432

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for causal mediation analysis. *Journal of Statistical Software*, *59*(5). https://doi.org/10.18637/jss.v059.i05

Van Hedger, S. C., Bongiovanni, N. R., Heald, S. L. M., & Nusbaum, H. C. (2023). Absolute pitch judgments of familiar melodies generalize across timbre and octave. *Memory & Cognition, 51*, 1898–1910. https://doi.org/10.3758/s13421-023-01429-z

Van Hedger, S. C., Heald, S. L. M., Koch, R., & Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition, 140*, 95–110. https://doi.org/10.1016/j.cognition.2015.03.012

Van Hedger, S. C., Heald, S. L., & Nusbaum, H. C. (2018). Long-term pitch memory for music recordings is related to auditory working memory precision. *Quarterly Journal of Experimental Psychology, 71*(4), 879–891. https://doi.org/10.1080/17470218.2017.1307427

Van Hedger, S. C., & Nusbaum, H. C. (2018). Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background. *Acta Psychologica, 191*, 251–260. https://doi.org/10.1016/j.actpsy.2018.10.007

Willander, J., & Baraldi, S. (2010). Development of a new Clarity of Auditory Imagery Scale. *Behavior Research Methods, 42*(3), 785–790. https://doi.org/10.3758/BRM.42.3.785

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory, 18*(6), 657–669. https://doi.org/10.1080/09658211.2010.501339

Xie, W., Park, H.-B., Zaghloul, K. A., & Zhang, W. (2020). Correlated individual differences in the estimated precision of working memory and long-term memory: Commentary on the study by Biderman, Luria, Teodorescu, Hajaj, and Goshen-Gottstein (2019). *Psychological Science, 31*(3), 345–348. https://doi.org/10.1177/0956797620903718