



What do less accurate singers remember? Pitch-matching ability and long-term memory for music

Andrea R. Halpern¹ · Peter Q. Pfordresher²

Accepted: 5 October 2021 / Published online: 18 November 2021
© The Psychonomic Society, Inc. 2021

Abstract

We have only a partial understanding of how people remember nonverbal information such as melodies. Although once learned, melodies can be retained well over long periods of time, remembering newly presented melodies is on average quite difficult. People vary considerably, however, in their level of success in both memory situations. Here, we examine a skill we anticipated would be correlated with memory for melodies: the ability to accurately reproduce pitches. Such a correlation would constitute evidence that melodic memory involves at least covert sensorimotor codes. Experiment 1 looked at episodic memory for new melodies among nonmusicians, both overall and with respect to the Vocal Memory Advantage (VMA): the superiority in remembering melodies presented as sung on a syllable compared to rendered on an instrument. Although we replicated the VMA, our prediction that better pitch matchers would have a larger VMA was not supported, although there was a modest correlation with memory for melodies presented in a piano timbre. Experiment 2 examined long-term memory for the starting pitch of familiar recorded music. Participants selected the starting note of familiar songs on a keyboard, without singing. Nevertheless, we found that better pitch-matchers were more accurate in reproducing the correct starting note. We conclude that sensorimotor coding may be used in storing and retrieving exact melodic information, but is not so useful during early encounters with melodies, as initial coding seems to involve more derived properties such as pitch contour and tonality.

Keywords Implicit/explicit memory · Music cognition · Sound recognition · Perceptual learning

Introduction

From the astonishingly accurate renditions of complex a cappella music sung by professional choirs to the inaccurate renderings of *Happy Birthday* sometimes heard at family celebrations, the range of pitch-matching accuracy in the population is both wide and obvious. Although occasional singers (i.e., people who do not sing in organized situations) may show a bias to underestimate their singing abilities (e.g., Pfordresher & Brown, 2007), objective tests of vocal matching for single pitches or short melodies have documented the wide range of performance in the population (Pfordresher & Demorest, 2021; Pfordresher & Larrouy-Maestri, 2015).

Because singing is a perceptual-motor skill, it provides an interesting model of how the auditory and motor systems interact (Berkowska & Dalla Bella, 2009; Hutchins & Moreno, 2013; Kleber & Zatorre, 2019; Loui, 2015; Pfordresher et al., 2015). Although inaccurate pitch matching could stem from a dysfunction in any of several processes, our focus is on the quality of the mental representation of the to-be-sung target, even among people with adequate auditory discrimination abilities and motor control. That internal representation is especially important in singing. In other motor behaviors, for instance playing a musical instrument, motor control can be refined by visual and tactile predictions and feedback. However, that control in singing is entirely covert, with no visual or tactile information to help predict, and adjust, the precise motor actions required to make fine-pitch changes.

Current models of the mechanisms underlying singing make predictions about the role of memory, but currently we have insufficient empirical data for validation of these predictions. The Vocal Sensorimotor Loop (VSL)

✉ Andrea R. Halpern
ahalpern@bucknell.edu

¹ Department of Psychology, Bucknell University, Lewisburg, PA 17837, USA

² Department of Psychology, University at Buffalo, SUNY, Buffalo, NY, USA

model proposed by Berkowska and Dalla Bella (2009), for instance, proposes bidirectional connections between memory (including long-term and working memory) and sensorimotor mapping. Given that most poor-pitch singing reflects deficits of sensorimotor mapping within the vocal system (e.g., Greenspon & Pfordresher, 2019; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007), this would suggest a link between singing accuracy and music memory. The Multi-Modal Imagery Association (MMIA) Model (Pfordresher et al., 2015) further elaborates on the nature of sensorimotor mapping, and proposes that the formation of auditory imagery, a type of perceptual memory, leads to multimodal associations with motor imagery (motor planning) that affect the accuracy and precision of vocal pitch matching, yet leaves the role of other kinds of memory unclear. This framework predicts an advantage for pitch matching of vocal timbres based on sensorimotor matching. This prediction has been borne out by previous studies (Hutchins & Peretz, 2012; Pfordresher & Mantell, 2014; Watts & Hall, 2008); however, it is important to note that learning to match pitch based on non-vocal timbres in the model (e.g., piano) is also possible.

Pfordresher and Halpern (2013) tested the relationship between pitch matching and auditory imagery in a large sample of undergraduate students. The percentage of correct single-note imitation (within a semitone) ranged from 100% in tune to 0%, with 11% singing in tune on none or only one of the six trials, 39% singing in tune on at least five of the six trials, and 22% singing in tune on all trials. We measured auditory imagery vividness by using a validated self-report scale: the Bucknell Auditory Imagery Scale or BAIS (Halpern, 2015). The BAIS has two subscales: Vividness, or BAIS-V, and Control, or BAIS-C, which captures reported ability to change one imagined sound into another. We found that BAIS-V correlated positively with pitch-matching accuracy for single pitches, suggesting not only a relationship between the skills but that people have reasonably good insight into the accuracy of their imagined sounds. We have also documented a relationship between imagery self-report, in this case BAIS-C, and accurate reproduction of more song-like short sound sequences (Greenspon et al., 2017).

Auditory imagery is of course a type of memory process. This memory representation can be activated from a long-term memory store, sometimes with great fidelity to the original stimulus. For instance, auditory imagery of familiar tunes represents pitch level and tempo remarkably consistently (Halpern, 1988; Jakubowski et al., 2015). Given our interest in individual differences in imagery, and the consequences of that for vocal production, we adopted that individual-differences approach here. Thus, we tested whether better pitch-matchers would also have enhanced long-term memory for music, presumably at least partly

due to enhanced auditory imagery skills. This would be a logical outcome of an elevated ability to generate, maintain, and consequently retrieve a representation of a wordless tune that does not have semantic referents.

We queried two domains of long-term memory. The first was incidental episodic learning of a list of new tunes, in particular, the Vocal Memory Advantage (VMA), which is the enhancement of episodic memory for melodies presented in a vocal compared to an instrumental timbre (Weiss et al., 2012). This superiority generalizes across variations in overall memory performance (Weiss et al., 2012), age (Weiss, Schellenberg, et al., 2015a), in people with otherwise deficient musical abilities (Weiss & Peretz, 2019), and among trained pianists (Weiss, Vanzella, et al., 2015b) – this latter result arguing against a simple familiarity explanation. Weiss and Peretz (2019) found that pitch-discrimination ability did predict overall memory for melodies (not surprisingly, as half their participants qualified as being amusic so had generally reduced pitch-processing abilities), but pitch discrimination did not predict magnitude of the VMA. Weiss and colleagues instead suggest that people are predisposed to attend more to human than nonhuman vocalizations, given the evolutionary importance of monitoring more than just the literal content of speech. That is, such a predisposition, one could argue, should also include the affective component of speech, which is often conveyed by prosody. In support of this suggestion, Weiss et al. (2016) found greater pupil dilation (a physiological measure of arousal) during listening to vocal than nonvocal tunes.

A logical consequence of enhanced attention is that it enables stronger memory representations via enhanced encoding. The hypothesis of main interest was that those with more robust sensorimotor coding, including (veridical) imagery representation, might be especially advantaged by the increased attention to vocalized melodies, leading to enhanced storage relative to piano melodies (that is, holding overall memory ability constant). In support of this hypothesis is the fact that motor programs are preconsciously activated when attending to sung versus played melodies. For instance, Lévêque and Schön (2015) demonstrated that the neural network associated with singing is more active when people were listening to human vocal timbres versus more artificial timbres. This attentional focus, linked to sensorimotor coding, would then result in a larger VMA among better pitch-matchers. A secondary hypothesis was that better pitch matchers might show overall better memory for melodies, if more accurate motor simulation was useful during encoding of any musical material. We left open the possibility that we would see this relationship more so for vocal than piano melodies.

The other domain we investigated was long-term memory for the absolute pitch of familiar recorded music, known as Absolute Pitch Memory (APM). APM

is an incidental type of learning accrued from exposure in everyday life. Even musically untrained people can show reasonably accurate memory for the pitch level of familiar songs. Levitin (1994) was the first to demonstrate higher-than-chance accuracy when people try to remember the first note of familiar recorded music, and this has more recently been replicated in a large multi-site study (although with a somewhat smaller effect size than originally reported; Frieler et al., 2013). The latter authors reported that fully a quarter of their sample of 277 people sang the precisely correct opening note for at least one of their two self-selected familiar songs. Some authors have documented APM in databases of folk music, and suggested APM is an aid in oral transmission and cultural memory of such songs (Olthof et al., 2015).

However, as with all results, there was considerable variability in success across participants in the replication study, with a standard deviation of nearly eight semitones (distance from the correct pitch). We located two prior studies that investigated some potential predictors of APM level. Jakubowski and Müllensiefen (2013) asked people to sing the first note of one popular recorded tune and found that a measure of relative pitch ability (verifying if two melodies were the same, after transposition of the second melody) predicted APM accuracy. Van Hedger et al. (2018) showed that some variability in APM can be accounted for by working memory span for pitch. In turn, we know that pitch WM is correlated with singing accuracy (Greenspon & Pfordresher, 2019), suggesting an association among the three factors of WM, singing accuracy, and APM. Interestingly, neither Jakubowski and Müllensiefen nor Van Hedger et al. found a contribution of musical training to APM accuracy.

Given these prior findings, we thought that APM might also be responsive to the quality of the imagery representation. A more veridical imagery representation would facilitate accurate singing when the person is singing along with the song on the radio, and, we propose, also when the song is called to mind without the song actually playing. When asked to produce the opening of a familiar song, participants typically try to first call the tune to mind, which we presume uses auditory imagery and motor simulation as a retrieval aid (the reader is invited to try this now). Accordingly, in Experiment 2, we asked people to select ten familiar songs recorded by only one artist in one key. To avoid confounds with variable skill in actual pitch production (i.e., the main predictor in this study), we asked them self-generate recognition candidates only by finding the correct starting note on a keyboard. The hypothesis was that people who could more accurately produce single notes and short melodies (better vocal pitch matchers) would have more skill and accuracy in the sensory to auditory imagery to motor simulation loop, which should enhance the implicit encoding of absolute pitch over repeated exposures in everyday life. Thus,

we predicted they would select notes closer to the actual recorded pitch.

In summary, we contrasted two types of memory for music, each of which could reasonably be inferred to benefit from superior functioning of a sensorimotor loop for pitch. Successful intentional learning of unfamiliar tunes after one presentation requires, among other things, attentional focus and the ability to quickly abstract pitch relationships. These also presumably are operating more efficiently in people more adept at singing back short patterns after one presentation, even though no singing or motor behavior is used in the recognition task. In contrast, APM represents long-term retention after multiple exposures (by definition, familiar songs are interrogated in such paradigms), but without conscious attempt to learn the actual key of the piece. However, the multiple exposures may have elicited and consolidated auditory-to-motor circuits, given the activity of motor systems during song listening referred to above (Lévêque & Schön, 2015) and the fact that pop music nearly always involves vocals. If that activity is reinstated during conscious retrieval of the opening line, even in the silent conditions imposed here, then people with better functioning sensorimotor loops normally used in imitation might be advantaged.

To foreshadow our results, using large samples showing a range of pitch-matching accuracy, we found support for the contribution of pitch-matching ability to long-term memory for absolute pitch, and for formation of new episodic memories for melodies, but no enhancement of the vocal memory advantage nor selectively better episodic memory for melodies in a vocal compared to a piano timbre.

Experiment 1

Methods

Participants The participants were 49 undergraduate students (14 male, 32 female, with the remainder not reporting gender) from Bucknell University aged 18–22 years. This sample size exceeds the size of $N = 40$ reported by Weiss and Peretz (2019) in their individual differences analysis and so was considered sufficient to detect significant associations with VMA. Years of musical training ranged from 0 to 12 (two participants did not report years of training), with a median of 1 year. Only one participant had more than 10 years of training. Participants received course credit as compensation.

Materials We used the same materials as Weiss et al. (2012). The set comprises 32 unfamiliar but tonal melodies derived from British and Irish folk songs, with durations ranging from 13 s to 19 s. For vocal melodies, each note was sung

on the syllable “la” by an amateur female alto singer, then edited for pitch and time accuracy. Piano melodies were played by an instrumentalist in the same tempo as the vocalist (facilitated by each vocal and instrumental performer hearing a backing track during recording).

Each melody was recorded in vocal and piano versions (leading to 64 total recordings), and different participants heard only one recording of each melody. Specifically, the 32 melodies were divided randomly into four sets of eight melodies. Participants heard two of the eight sets during the encoding phase and all four sets during the test phase, with half the sets in each phase being in a piano and half in a vocal timbre. Each of those sets was rotated such that each melody was presented equally often across participants as (a) an old or a new melody and (b) in the vocal or the piano timbre.

Singing accuracy for pitch was assessed through the Seattle Singing Accuracy Protocol (SSAP; Demorest & Pfordresher, 2015; Demorest et al., 2015; Pfordresher & Demorest, 2020), an on-line measure. The relevant section of the SSAP is a series of imitative (call-and-response) singing trials in which participants match single-note targets (in both vocal and piano timbre) and short melodies (in a vocal timbre), presented in each person’s comfortable vocal range. The SSAP includes an automated scoring procedure (for details see Pfordresher & Demorest, 2020), in which pitches sung within ± 50 cents of the target pitch (one semitone total range) are scored as accurate. This procedure involves the extraction of vocal f_0 using the Matlab function YIN (de Cheveigné & Kawahara, 2002), with pitch for individual notes estimated using the median of the middle portion of each note. Note segmentation for short melodies is accomplished using peaks in the amplitude contour associated with the syllable “doo” used for each note. An additional musical background questionnaire collected information about years of training and musical participation, as well as demographic data.

Participants’ self-reported vividness and control of auditory imagery was assessed via the Bucknell Auditory Imagery Scale (BAIS, Halpern, 2015). This measure includes two subsets of items designed to measure self-reported vividness of auditory images based on music, speech, and environmental sounds generated from long-term memory (e.g., the sound of gentle rain), as well as the ease with which participants can modify auditory images (e.g., rain becomes a thunderstorm).

Procedure Participants were run individually in a quiet, isolated room. After the consent form was signed, the encoding phase of the memory task followed the procedure of Weiss et al. (2012). After instructions were presented on a screen and participants given a chance to answer any questions, each of the 16 to-be-remembered melodies (eight piano and eight vocal), was played in a random order over high-quality speakers at a comfortable listening level on a MacBook Pro computer, running SuperLab software. To ensure attention to each melody,

an incidental encoding task involved asking participants to rate each melody as sounding happy (1), neutral (2), or sad (3) by pressing the number corresponding to the intended emotion judgment. No mention was made of a later recognition phase.

In the encoding phase, the presentation of each melody followed this sequence: first, participants were shown a 3-s visual countdown (comprising the inter-stimulus interval), followed by the melody presented along with a visual symbol (a musical note). After the melody ended, participants made the emotion judgment. Then, participants took the BAIS on paper.

In the recognition phase, 32 melodies were presented one at a time in random order: 16 in the vocal timbre and 16 in the piano timbre. Half of each melody type had been presented in encoding phase, and half were new melodies. Participants were asked to rate each tune on a paper response form on a 7-point scale, with 1 denoting *definitely new* and 7 denoting *definitely old*. Participants were encouraged to use the full range of the scale in their responses.

Following the recognition phase, participants were administered the SSAP, followed by the Musical Background questionnaire. Participants were then thanked and given a debriefing statement.

Results

Vocal memory advantage As a first step, we assessed whether the recognition data from Experiment 1 replicated the Vocal Memory Advantage (VMA) reported by Weiss et al. (2012). Following their protocol, recognition accuracy was based on the difference between the mean rating score given to melodies not encountered during the exposure phase, which should yield a rating of “1” for a high confidence correct rejection, from the mean rating for melodies that participants had heard before, where “7” would be a high confidence hit. For each participant, the mean rating for all new melodies was subtracted from the mean rating for all previously exposed melodies. Perfect responding across all melodies would yield a difference score of 6, with chance represented by a difference score equal to 0.

Participants were well above chance for recognizing melodies presented in both timbres; however, recognition was significantly more accurate (i.e., closer to a score of 6) for melodies presented in a vocal timbre (M difference = 2.46, $SE = 0.19$) than melodies presented in a piano timbre (M difference = 1.88, $SE = 0.16$), $t(48) = 3.70$, $p < .001$. For each participant, the predictor was the VMA score based on subtracting mean accuracy across all recognition trials for piano melodies, from the mean accuracy across all recognition trials for vocal melodies. Positive values indicate a Vocal Memory Advantage, whereas negative values indicate an advantage for the piano timbre. As indicated by the above, mean VMA was .58 ($SE = .16$). We thus replicated the Vocal

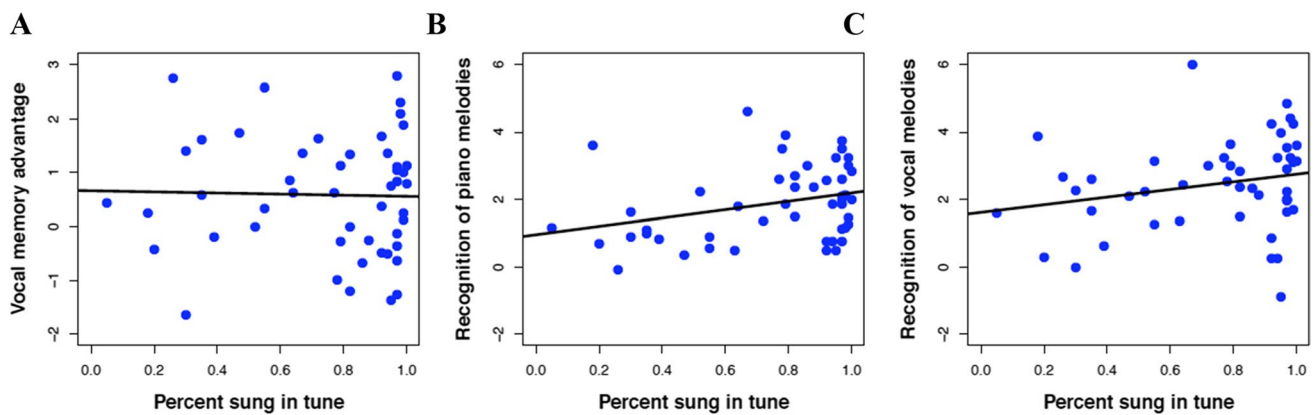


Fig. 1 Mean percent of pitches matched within ± 50 cents across all imitation trials of the SSAP (X, all panels) by mean difference between recognition performance for melodies presented in a vocal timbre minus melodies presented in a piano timbre (Y, **panel A**),

recognition performance for melodies presented in a piano timbre (**B**), and recognition performance presented in a vocal timbre (**C**). Each data point represents the mean for a single participant. The solid line represents best-fitting least squares regression

Memory Advantage overall in this sample of participants (which, we note, had on average less musical training than the original sample, most being completely untrained). We also replicated the lack of correlation between overall memory and the VMA reported by Weiss and Peretz (2019), here $r(47) = .24, p = .10$, reinforcing the differentiation of the relative measure of the VMA from absolute memory score.

Singing accuracy and correlations Singing accuracy was measured using the mean proportion of pitches sung correctly during the Seattle Singing Accuracy Protocol (SSAP). Produced pitches were coded as errors if the absolute difference between the fundamental frequency (f_0) of the sung pitch and the f_0 of the target pitch exceeded 50 cents (half the distance between two pitches in the equal tempered tuning system). These errors were aggregated across all notes and trials for the three pitch imitation tasks within the SSAP. Subsequent analyses based on different measures of accuracy and separate subtests of the SSAP yielded similar results to those reported here.¹ Figure 1A plots the relationship between the mean proportion of correct pitches for each participant and VMA scores. The relationship was not statistically significant and slightly negative, $r(47) = -.03, p = .86$. This lack of correlation held after both arcsin and logit transformations, applied due to the nonnormality of the distribution. Associations between singing accuracy and recognition accuracy (difference in ratings for exposed versus unexposed melodies) were stronger and positive, with a significant association with recognition of melodies averaged over piano and vocal ($r(47) = .30, p = .04$), and also just for items presented in a piano timbre, $r(47) = .31, p = .03$

(Fig. 1B). A similar pattern but not statistically robust was seen with recognition of melodies presented in a vocal timbre, $r(47) = .23, p = .10$ (Fig. 1C); however, these two correlations did not differ significantly when directly compared (Fisher's r to z transformation; $z = .644$). Correlations of the memory scores with BAIS subtests were non-significant, although there was a marginal correlation of BAIS-V with memory for piano timbre melodies ($r(47) = .28, p = .05$).

Discussion

We replicated the VMA, and its independence from absolute memory performance, adding to the data on robustness of this finding. However, we found no association between this vocal advantage and singing accuracy. We did find a correlation between singing accuracy and memory, although interestingly, not more strongly for vocal compared to piano melodies (with a trend towards a stronger association for piano compared to vocal), failing to also support our secondary hypothesis. Hence, we conclude that encoding of a sequence that could *in theory* be implemented by covertly activating a vocal code or singing simulation does not in fact effectively make use of the sensorimotor coding used in pitch replication, at least in silent conditions. This is also supported by the lack of correlation of the VMA (or memory for vocal melodies) with self-rated auditory imagery vividness or control.

Experiment 2

Method

Participants The participants were 46 undergraduate students (14 male) from Bucknell University aged 18–22 years.

¹ Singing accuracy data from one participant was lost and so this participant was removed from these analyses. All correlations are one-tailed tests due to strong a priori predictions about direction of relationship.

Years of musical training ranged from 0 to 12 years with a median of 2.00. This sample size is comparable to the $N = 50$ reported by the Jakubowski and Müllensiefen (2013) study and exceeds the sample sizes of $N = 29$ and 40 in the two experiments by Van Hedger et al. (2018). Only one participant had more than 10 years of musical training. Participants received course credit as compensation.

Materials Participants made selections from a master list of 47 songs that had to meet several criteria: each could only have been recorded by one artist (and thus in one key), be very familiar to our cohort of university students, and have a distinct opening note. For the experiment proper, each person selected the ten songs that were most familiar, for which the opening note would be retrieved by playing it on a keyboard

We consulted the 2016 *Billboard Top 100* list as well as *The Rolling Stone Top 500* songs for initial candidates (these data were collected in 2017, so were getting airtime during that period). From that list, we eliminated songs that had been covered by other musicians, verified by checking several databases of song recordings. Songs were also eliminated that did not meet the other listed criteria. Familiarity was initially assessed by the undergraduate research assistants working on this project. For each song passing that stage, we located the recording on Spotify or iTunes, and a trained musician assessed whether there was an unambiguous opening note, and, if so, labeled it with the standard Western note name (usually this was the main vocal line but occasionally a distinctive instrumental opening). This phase resulted in a set of 53 candidate songs.

To verify familiarity for these songs, we presented the titles, plus the artist and opening lyrics of each to a set of seven naïve undergraduates. They were asked to rate each song on a familiarity scale where 1 meant “not familiar” and 5 meant “so familiar I can ‘hear’ the song in my head.” We set an inclusion criterion of mean rating 4.5 to be included in the master list. A total of 47 songs were kept on the master list after this phase.

Procedure After signing the consent form, participants filled out a musical background questionnaire. They then saw a computer screen on which they could choose the ten songs they were most familiar with from the master list (participants were allowed to continue in the study if they knew at least eight songs). The experimental software then pulled just those trials for presentation in the main phase. Appendix 1 lists song candidates and the frequency with which each was selected.

The participants were seated near an electronic keyboard and the experimenter read the instructions for the absolute pitch memory task:

The tunes you selected will appear on the screen one at a time with the artist and a lyric from the song. Given the lyric, think of the pitch that line begins on in the recorded version of the song. Please do not sing or hum it; just

think it. Please try to find that first pitch of the given line on the keyboard. We will be looking at how long it takes you, but we are most interested in your being accurate. When you think you have found the correct starting pitch please hold that key down as the last selection for that song, then look up to let me know you’re finished.

Each of the ten selected songs was then presented once, in random order. The participant initiated the trial. After the participant indicated satisfaction with the selected note, the experimenter recorded the note name and the participant initiated the next trial. The experimenter remained in the room with the participant during the session and monitored each trial to ensure that no overt vocalization took place during note-finding.

After this task was completed, participants took the BAIS. A short break then ensued after which participants filled out the musical background questionnaire and took the SSAP. Participants were then thanked and debriefed.

Results

Absolute pitch memory Figure 2 shows the distribution of scores computed as the signed difference between participants’ selected starting pitch and the actual starting pitch of the song, averaged over songs. Because responses here were not constrained by vocal range, we treated starting pitches produced in the wrong octave as errors, whereas Levitin (1994) analyzed errors in an octave-equivalent manner. The most common mean deviation score was centered around zero, with 17% of participants generating means within ± 1 semitone and 33% participants within ± 2 semitones. Although this performance falls somewhat below the rates of accuracy reported by Levitin (1994), in which 44% of participants were accurate to within two semitones, Levitin had participants generate starting pitches by singing; finding the pitch on a piano keyboard is likely a more difficult output task (and as above, we did not equate octaves). It is worth noting that this accuracy of absolute pitch memory still exceeds by several orders of magnitude the ability of ordinary people (non-possessors of absolute pitch, or what is sometimes referred to as “perfect pitch”) to generate the correct note names for pitches presented in isolation (Levitin & Rogers, 2005).

For the purposes of correlating absolute pitch memory with singing accuracy, we were not concerned with the direction of pitch errors. Therefore, the mean unsigned deviation scores for participants reflect the overall accuracy of starting pitches selected from the keyboard.

Correlations with singing accuracy and other measures Singing accuracy was measured using the same procedure as in Experiment 1. Mean percent correctly sung pitches did not differ significantly across the two experiments (Experiment

1: $M = 77\%$, $SD = 27\%$; Experiment 2: $M = 72\%$, $SD = 22\%$; $p = .52$). However, unlike Experiment 1, in Experiment 2 singing accuracy correlated significantly with mean absolute deviations in pitch memory, $r(44) = -.30$, $p = .02$, as shown in Fig. 3. Participants who more often imitated pitches within one semitone of the target also exhibited more accuracy when selecting the starting pitch of a familiar melody. This correlation held after both arcsin square root ($r = -.30$, $p = .02$) and logit ($r = -.29$, $p = .03$) transformations, applied due to the nonnormality of the distribution. The BAIS-V did not correlate with mean unsigned error in the APM task ($r(44) = .02$, $p = .56$), but it did correlate with mean pitch accuracy in the SSAP, $r(44) = .26$, $p = .04$. BAIS-C did not correlate with either outcome measure, $r = .14$ in both cases. Years of musical training likewise was not correlated with mean unsigned error in the APM task, $r(44) = -.24$, $p = .12$.

Discussion

Our goal was to examine the relationship between two processes that on the surface do not appear to be highly related: the sensorimotor system involved in matching pitch in singing, and the cognitive skill of remembering music. We had reason to propose that better pitch matchers would use that skill, possibly involving auditory imagery, to attend to, encode, and retrieve melodies. Considering first the challenging task of recognizing a list of unfamiliar tunes after one exposure, although we replicated the Vocal Memory Advantage (Weiss et al., 2012), in that people did remember sung melodies better than piano-rendered melodies, surprisingly and contrary to our major hypothesis, the better pitch matchers did not show a larger VMA. If we make the reasonable assumption that better pitch matchers would have a more accurate representation of tunes they themselves could have produced (i.e., pitch accuracy maintained in the cycle of hearing-encoding-reproducing-encoding-reproducing, etc.), it seems reasonable to predict they would covertly encode sung stimuli using that same loop, with success in one task associated with success in the other. That not being the case, our result supports the idea proposed by Weiss et al. (2016) that the VMA is primarily a consequence of initial attentional rather than downstream memory encoding processes.

Regarding our secondary hypothesis, pitch accuracy was not more associated for the vocal compared to the piano melodies (with a trend in the reverse direction). The trend to a more robust relationship between pitch matching skill and piano tunes is consistent with a general factor of good auditory processing. As hearing a live or recorded melody in piano is likely a more common experience than hearing a person singing “la-la-la,” practice in encoding piano melodies is likely more common and perhaps overall better auditory processing better pitch matchers.

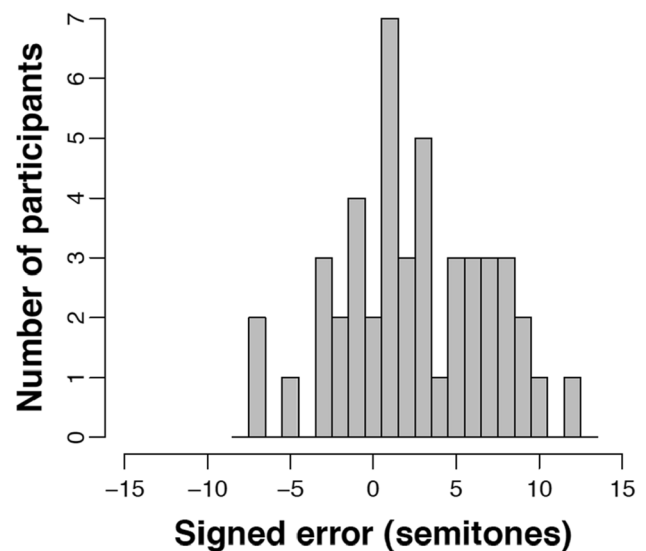


Fig. 2 Histogram reflecting the mean difference between the starting pitch chosen by each participant, minus the starting pitch of the song for each participant, across all songs selected. Deviations are in semitones, with positive values indicating that the participant chose a pitch that was too high on average

Recently, a debate has surfaced concerning whether the VMA is based on the use of subvocalization during the encoding of melodies, which may more effectively mimic the structure of vocal melodies than melodies presented in another timbre. Two recent studies used interference paradigms to test this view, yet led to opposite findings, with one study mentioned above showing disruption of VMA during rapid repetition of a whispered syllable (Wood et al., 2020), and another finding no such disruption for various articulatory interference tasks (Weiss et al., 2020). The present study bears on this issue insofar as poor-pitch singers are thought to suffer from inadequate sensorimotor associations that are correlated with measurable subvocalization in muscles used for the control of pitch (Pruitt et al., 2019). In this context, the present results once again comport with the view that the vocal memory advantage is based on voice-specific initial processing and not on subvocal reproduction (as per Weiss et al., 2020), given that no reproduction tasks were used here. However, our correlational design does not provide as powerful a test as would an interference paradigm, which can be used to assess if subvocalization is necessary (essential) for the VMA.

In Experiment 2, we explored longer-term, but incidental memory of absolute pitch for familiar melodies. Similar to prior studies, we saw a fairly normal distribution of accuracy in selecting that note from a keyboard. That is, contrary perhaps to what most non-possessors of traditional Absolute Pitch would predict, a substantial number of our participants were spot-on or close to the correct opening pitch, including the correct octave. And this, we saw, was in fact correlated with accuracy in the very different task of vocal matching single

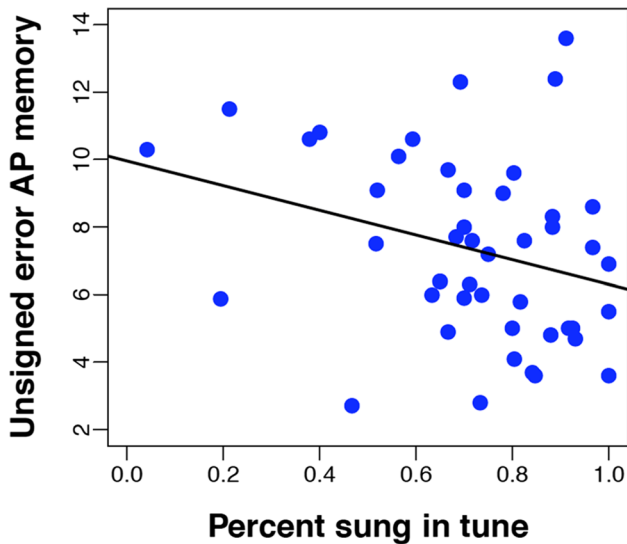


Fig. 3 Mean percent of pitches matched within ± 50 cents across all imitation trials of the SSAP (X) by mean absolute difference between the starting pitch chosen by each participant, minus the starting pitch of the song for each participant across all songs selected (Y). Each data point represents the mean for a single participant. The solid line represents the best-fitting least squares regression

pitches or short patterns. This suggests that the sensorimotor patterns involved in pitch matching can be co-opted for long-term memory encoding of items that do not have ready semantic representations (i.e., only people with absolute pitch would be able to verbally encode the name of the opening note and produce the note from that label). This interpretation would be consistent with a model of vocal perception and production proposed by Hutchins and Moreno (2013) called the Linked Dual Representation model, wherein vocal-motor production schemas can operate even in the absence of conscious knowledge about tones such as is used in labeling and categorization decisions. It is worth noting that the opening note of some of the songs involved instrumental introductions, so the pitch learning need not necessarily have come from actually singing a vocal line along with a recording. And, of course, no overt singing was allowed during the task.

Based on our prior work, we had proposed that auditory imagery was a specific candidate mechanism for facilitating better memory in better pitch matchers. And in Experiment 2, we replicated our usual finding of the BAIS-V (but not BAIS-C) predicting singing error. However, in these samples, we did not find a significant contribution to recognition memory, the VMA, or APM from the BAIS self-report scale. It would be interesting to see if objective measures of auditory imagery ability would have predictive power (cf. Greenspon & Pfordresher, 2019). Another possibility is that moment-to-moment (state) rather than trait measures of imagery would capture this mediating variable. For instance, Pruitt et al. (2019) found that ratings of trial-by-trial vividness (state)

predicted performance in a singing imitation task, rather than the BAIS (trait). However, the current results do at least implicate the relationship of embodied skill of pitch matching with the sensory-like experience of recalling an (unnamed) absolute pitch, reinforcing the importance of integrated sensory-cognitive-motor interactions in auditory cognition.

These results have important theoretical implications for the relationship between memory and the mechanisms underlying singing. Figure 4 illustrates the relationship between the present results and the VSL model of Berkowska and Dalla Bella (2009), adapted to include details of sensorimotor mapping from the MMIA model (Pfordresher et al., 2015). We propose that that the Vocal Memory Advantage (VMA) reflects attentional weighting of auditory inputs to memory, based both on our data and previous studies of VMA (Weiss et al., 2016). This connection in the VSL model is independent of sensorimotor mapping and is thus dissociable from most individual variance in singing accuracy. By contrast, our data suggest that the retrieval of long-term AP memory may involve connections between memory and the imagery processes that underly sensorimotor translation, thereby exhibiting associations with singing accuracy. Specifically, the need to retrieve specific pitch information from long-term memory may draw more heavily on the formation of a multimodal image than episodic recognition. These proposals are tentative given that they are based on two particular instantiations of the memory paradigm. We see the present results, and their theoretical implications, as making an important step forward in understanding the links between different types of memory and sensorimotor functions involved in music performance.

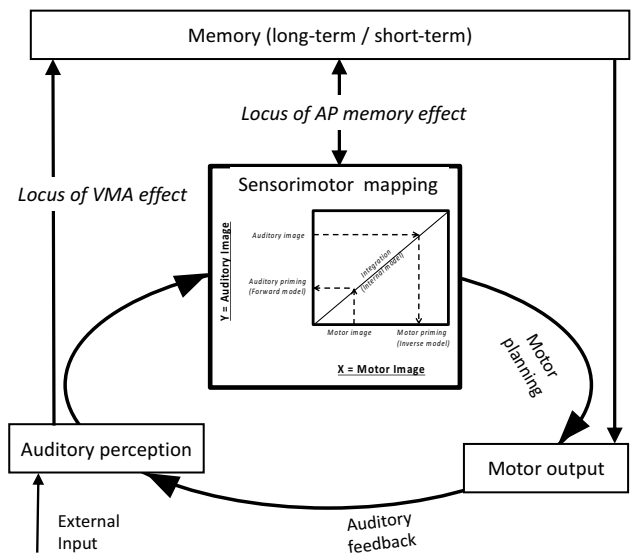


Fig. 4 Model of vocal pitch matching including hypothetical associations with memory based on the present results. The model here is derived from a combination of the Vocal Sensorimotor Loop model (Berkowska & Dalla Bella, 2009), with an elaboration of sensorimotor mapping based on the MMIA model (Pfordresher et al., 2015)

Appendix

Table 1

Title	Artist	Frequency
Hello	Adele	28
Closer	The Chainsmokers	28
Sorry	Justin Bieber	26
Love Yourself	Justin Bieber	20
Shape of You	Ed Sheeran	20
My House	Flo Rida	20
One Dance	Drake	17
iSpy	Lil Yachty	17
Hotline Bling	Drake	17
Work	Rihanna	16
That's What I Like	Bruno Mars	12
Treat You Better	Shawn Mendes	12
Paris	The Chainsmokers	11
Mask Off	Future	11
Castle on a Hill	Ed Sheeran	11
Something Just Like This	The Chainsmokers and Coldplay	10
Starboy	The Weeknd	10
Stitches	Shawn Mendes	10
Stressed Out	21 Pilots	9
Stay	Zedd and Alessia Cara	9
Mercy	Shawn Mendes	9
Roses	The Chainsmokers	9
Rolox	Ayo and Teo	8
Cake by the Ocean	DNCE	8
Cold Water	Major Lazor	8
Love on the Brain	Rihanna	7
Scars To Your Beautiful	Alessia Cara	7
Million Reasons	Lady Gaga	7
Cheap Thrills	Sia and Sean Paul	7
This is What You Came For	Calvin Harris ft. Rihanna	7
Me, Myself, and I	G-Eazy	7
Too Good	Drake	7
I Feel it Coming	The Weeknd and Daft Punk	6
I Don't Wanna Live Forever	Zayn and Taylor Swift	6
Tunnel Vision	Kodak Black	6
Animals	Maroon 5	6
Heathens	21 Pilots	6
Can't Stop the Feeling	Justin Timberlake	5
Don't Let Me Down	The Chainsmokers ft. Daya	5
Ride	21 Pilots	5
The Hills	The Weeknd	5
Passionfruit	Drake	4
Rockabye	Clean Bandit ft Sean Paul	4
Swalla	Jason Derulo	4
Work From Home	Fifth Harmony	3
Pillowtalk	Zayn	0
Just Like Fire	P!nk	0

Acknowledgements We thank Michael Weiss for providing the melodies used in Experiment 1. We also thank research assistants Matthew DiSanto, Nura Gouda, Brent Mankin, Brendan Matthys, Siobhan Nerz, Stephanie Purnell, and Samantha Santomartino.

References

- Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69-83.
- de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, *111*, 1917-1931.
- Demorest, S. M. & Pfordresher, P. Q. (2015). *Seattle Singing Accuracy Protocol – SSAP* [Measurement instrument]. <https://ssap.music.northwestern.edu/>
- Demorest, S. M., Pfordresher, P. Q., Bella, S. D., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: an introduction to the special issue on singing accuracy (Part 2). *Music Perception: An Interdisciplinary Journal*, *32*, 266-271.
- 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*, 334-349.
- Greenspon, E. B., & Pfordresher, P. Q. (2019). Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation. *Attention, Perception, & Psychophysics*, *81*, 2473-2481.
- Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation in mental transformations of melodies. *Music Perception*, *34*, 585- 604.
- Halpern, A. R. (1988). Perceived and imagined tempos of familiar songs. *Music Perception*, *6*, 193-202.
- Halpern, A. R. (2015). Differences in auditory imagery self-report predict behavioral and neural outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*, 37-47.
- Hutchins, S., & Moreno, S. (2013). The linked dual representation model of vocal perception and production. *Frontiers in Psychology*, *4*, 825.
- Hutchins, S., & 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*, 76-97.
- Jakubowski, K., Farrugia, N., Halpern, A.R., Sankardpandi, S. & Stewart, L. (2015). The speed of our mental soundtracks: Tracking the tempo of involuntary musical imagery in everyday life. *Memory and Cognition*, *43*, 1229-1242.
- 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*, 1259-1267.
- Kleber, B., & Zatorre, R. (2019). The neuroscience of singing. In G. F. Welch & J. Nix (Eds.), *The Oxford Handbook of Singing*. : Oxford University Press.
- Lévêque, Y., & Schön, D. (2015). Modulation of the motor cortex during singing-voice perception. *Neuropsychologia*, *70*, 58-63.
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*, 414-423.
- Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*, 26-33.
- Loui, P. (2015). A dual-stream neuroanatomy of singing. *Music Perception*, *32*, 232-241.
- Olthof, M., Janssen, B., & Honing, H. (2015). The role of absolute pitch memory in the oral transmission of folksongs. *Empirical Musicology Review*, *10*(3), 161-174.
- Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception*, *25*, 95-115.
- Pfordresher, P. Q., & Demorest, S. M. (2021). The prevalence and correlates of accurate singing. *Journal of Research in Music Education*, *69*, 5-23.

- Pfordresher, P. Q., & Demorest, S. M. (2020). Construction and validation of the Seattle Singing Accuracy Protocol (SSAP): An automated online measure of singing accuracy. In F. Russo, B. Ilari, & A. Cohen (Eds). *Routledge Companion to Interdisciplinary Studies in Singing: Vol. 1 Development* (pp. 322-333). : Routledge.
- Pfordresher, P. Q. & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin and Review*, 20, 747-753.
- Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The Multi-Modal Imagery Association (MMIA) Model. *Music Perception*, 32, 293-302.
- Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we understand deficits of vocal pitch imitation? *Frontiers in Human Neuroscience*, 9, 27.
- Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology*, 70 31-57.
- Pruitt, T. A., Halpern, A. R., & Pfordresher, P. Q. (2019). Covert singing in anticipatory auditory imagery. *Psychophysiology*, 56(3), e13297.
- 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*, 74, 879-891.
- Watts, C. R., & Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. *Logopedics Phoniatrics Vocology*, 33, 74-82.
- Weiss, M. W., Bissonnette, A.-M., & Peretz, I. (2020). The singing voice is special: Persistence of superior memory for vocal melodies despite vocal-motor distractions. *Cognition*, 104514.
- Weiss, M. W., & Peretz, I. (2019). Ability to process musical pitch is unrelated to the memory advantage for vocal music. *Brain and Cognition*, 129, 35-39.
- Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, 23, 1074-1078.
- Weiss, M. W., Schellenberg, E. G., Trehub, S. E., & Dawber, E. J. (2015a). Enhanced processing of vocal melodies in childhood. *Developmental Psychology*, 51, 370.
- Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology: Human Perception and Performance*, 42, 1061.
- Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015b). Rapid Communication: Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *Quarterly Journal of Experimental Psychology*, 68, 866–877.
- Wood, E. A., Rovetti, J., & Russo, F. A. (2020). Vocal-motor interference eliminates the memory advantage for vocal melodies. *Brain and Cognition*, 145, 105622.

Open Practices

Neither of the experiments was preregistered. Stimuli for Experiment 1 were provided by Michael Weiss; only names of songs were used as materials in Experiment 2 and are provided in the Appendix. Data from both experiments are available upon request from the first author.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.