Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–13

DOI: 10.1080/17470218.2017.1307427
qjep@sagepub.com
SAGE

Long-term pitch memory for music recordings is related to auditory working memory precision

Stephen C Van Hedger, Shannon LM Heald and Howard C Nusbaum

Abstract

Most individuals have reliable long-term memories for the pitch of familiar music recordings. This pitch memory (1) appears to be normally distributed in the population, (2) does not depend on explicit musical training and (3) only seems to be weakly related to differences in listening frequency estimates. The present experiment was designed to assess whether individual differences in auditory working memory could explain variance in long-term pitch memory for music recordings. In Experiment 1, participants first completed a musical note adjustment task that has been previously used to assess working memory of musical pitch. Afterward, participants were asked to judge the pitch of well-known music recordings, which either had or had not been shifted in pitch. We found that performance on the pitch working memory task was significantly related to performance in the pitch memory task using well-known recordings, even when controlling for overall musical experience and familiarity with each recording. In Experiment 2, we replicated these findings in a separate group of participants while additionally controlling for fluid intelligence and non-pitch-based components of auditory working memory. In Experiment 3, we demonstrated that participants could not accurately judge the pitch of unfamiliar recordings, suggesting that our method of pitch shifting did not result in unwanted acoustic cues that could have aided participants in Experiments 1 and 2. These results, taken together, suggest that the ability to maintain pitch information in working memory might lead to more accurate long-term pitch memory.

Keywords

Working memory; implicit memory; individual differences; music perception; absolute pitch

Received: 4 January 2016; revised: 4 January 2017; accepted: 9 January 2017

Absolute pitch (AP) is typically defined as the rare ability to name or produce a musical note without the aid of a reference note (see Takeuchi & Hulse, 1993 for a review). This ability is supported by explicit, long-term category structures for isolated musical notes, which the vast majority of listeners are not thought to possess. In some sense, this could be taken to suggest that without such long-term category structures, the auditory trace of a sound is not preserved outside of auditory working memory (cf. Rakowski & Morawska-Bungeler, 1987). However, recent research has argued for a ubiquitous form of long-term pitch memory even in those without AP, sometimes referred to as *implicit, residual or latent AP*. This kind of pitch memory does not depend on explicit pitch labels and manifests in several ways. For example, individuals are able to produce songs with very little AP variance, even across different experimental sessions (Halpern, 1989). Individuals are also sensitive to the absolute tuning of popular songs in their environment, as evidenced by the fact that when they are asked to produce well-known melodies (through humming, singing or whistling), they choose the correct key signature at a rate that is higher than chance (Frieler et al., 2013; Levitin, 1994). Beyond production, individuals can reliably distinguish when well-known music recordings have been shifted in pitch, even if the shift is as subtle as one semitone—the smallest pitch difference found in conventional Western music (Schellenberg & Trehub, 2003). Even when using recently learned, experimentally controlled melodies,

Department of Psychology, The University of Chicago, Chicago, IL, USA

Corresponding author:
Stephen C Van Hedger, Department of Psychology, The University of Chicago, 5848 S. University Ave., Chicago, IL 60637, USA.
Email: shedger@uchicago.edu

adult listeners have been shown to be sensitive to both absolute and relative pitch cues (Creel & Tumlin, 2012) even after a single exposure (Schellenberg & Habashi, 2015; Schellenberg, Stalinski, & Marks, 2014). Beyond music recordings, individuals display remarkably accurate pitch memories for environmental sounds, such as the landline dial tone (Smith & Schmuckler, 2008) and isolated tones such as the "bleep" used to censor taboo words in the media (Van Hedger, Heald, & Nusbaum, 2016). These findings strongly suggest that good long-term pitch memory exists beyond the boundaries of genuine AP.

What factors influence the strength of this pitch memory? It is clear that familiarity plays a foundational role as individuals are not able to determine the correct or incorrect tunings of unfamiliar melodies (Schellenberg & Trehub, 2003). Beyond specific auditory objects heard in the environment (e.g., recordings of melodies and dial tone), recent research has demonstrated that everyday listeners are sensitive to the statistical regularities of note intonation in their musical environments and can apply this knowledge to label an isolated note as "in-tune" or "out-of-tune" (Van Hedger, Heald, Huang, Rutstein, & Nusbaum, 2017). Furthermore, everyday listeners are sensitive to how often individual notes are heard, in that less frequently experienced notes are more preferred to more frequently experienced notes, even when presented in isolation (Ben-Haim, Eitan, & Chajut, 2014). This role of environmental experience in the establishment of preference is by no means specific to music. For example, expert typists perceptually prefer letter dyads that would contain less motor interference (e.g., "DJ" vs "DV") if they were to be typed, even though the typists were not aware of this rule and were not physically typing the letter dyads (Beilock & Holt, 2007).

Despite the important role of environmental experience in the grounding of implicit pitch memory, listener-produced subjective estimates of how often one has heard a particular recording do not strongly correlate with pitch memory for popular recordings (Schellenberg & Trehub, 2003). While this weak relationship could be due to inaccurate self-reports of how frequently one has listened to particular recordings, it could also be the case that environmental experience is important only insofar as it initially establishes pitch memory (i.e., beyond a certain point, more exposure does not necessarily correspond with a more accurate pitch memory). Indeed, more recent work suggests that a single listening experience is enough to instantiate AP memory for novel music recordings (Schellenberg & Habashi, 2015; Schellenberg, Stalinski, & Marks, 2014). This suggests that it is one's *familiarity* with a particular recording and not one's accrued listening time for a given recording that appears to predict pitch memory performance.

Another factor that, on the surface, might appear to be related to pitch memory strength is explicit musical training. Indeed, musicians have been shown to have smaller difference limens for auditory pitch (Kishon-Rabin, Amir, Vexler, & Zaltz, 2001), better verbal memory abilities (Chan, Ho, & Cheung, 1998; Ho, Cheung, & Chan, 2003), better cognitive control of auditory working memory (Pallesen et al., 2010) and better speech-in-noise detection compared to non-musicians (Parbery-Clark, Skoe, Lam, & Kraus, 2009). Despite these musician advantages, which extend beyond musical processing to general auditory processing, it does not appear to be the case that explicit formal musical training strongly predicts accuracy in pitch memory (Jakubowski & Müllensiefen, 2013; Levitin, 1994; Schellenberg & Trehub, 2003; though see Frieler et al., 2013). However, one component of explicit musical training that may affect pitch memory accuracy is relative pitch memory, which has recently been found to relate to pitch memory for familiar recordings (e.g., Bartlette, Henry, & Moore, 2015; Jakubowski & Müllensiefen, 2013).

Given that both song familiarity and musical training do not appear to strongly relate to the individual differences observed in pitch memory, this study aims to address whether a previously unexplored factor—auditory working memory—explains any individual variance in long-term pitch memory. There are two primary reasons to suspect that auditory working memory might relate to long-term pitch memory. First, recent work has shown that auditory working memory ability predicts how well adults can explicitly acquire AP categories, even when controlling for the age at which individuals began musical instruction (Van Hedger, Heald, Koch, & Nusbaum, 2015). While this relationship between auditory working memory and AP was discussed in terms of explicit AP category acquisition, it could be the case that individuals who are able to maintain more accurate auditory representations in working memory would also have more accurate implicit long-term pitch representations for familiar recordings. Second, Jakubowski and Müllensiefen (2013) found that *relative pitch* perception was related to pitch accuracy in producing familiar melodies. However, relative pitch perception was measured by having participants listen to two melodies and determining whether or not they were identical. While this task certainly requires relative pitch processing, it also requires holding onto the first melody in working memory in order to make an accurate comparison with the second melody.

The present experiments, therefore, investigate the relationship between auditory working memory and long-term pitch memory for musical recordings while controlling for several additional factors (e.g., musical training and recording familiarity). In Experiment 1, we assess auditory working memory through an implicit note memory (INM) procedure, in which participants had to adjust a starting note to match a previously heard target note. This procedure has been previously demonstrated to share variance with more canonical measures of auditory working memory (specifically, the auditory n-back), and it has also been shown to explain variance in explicit AP category acquisition (Van Hedger et al., 2015). In Experiment 2, we assess whether the relationship between long-term pitch memory and auditory working memory exists within non-musical measures of auditory working memory and short-term memory (using forward digit span and auditory n-back), as well as within more general measures of fluid intelligence (using Raven's Advanced Progressive Matrices [RAPM]). These are important considerations for understanding (1) the relationship between more general measures of auditory working memory and long-term pitch memory and (2) assessing whether domain-general differences in executive attention or "cognitive fitness" can similarly predict individual differences in pitch memory. In Experiment 3, we assess whether the long-term pitch memory accuracy observed in Experiments 1 and 2 could actually be explained by artifacts in our pitch-shifting procedure, rather than true long-term memory for musical recordings.

Experiment 1

Method

Participants. In total, 29 University of Chicago undergraduates participated in the study for course credit (M_{age} = 19.90 years, SD_{age} = 1.70 years, range = 18-27 years, 18 females, 11 males). Participants were not specifically recruited for their musical experience, although a majority of participants (25 of 29) reported at least some music instruction. Further details of participants' musical training are reported in section "Results." No participant reported possessing AP.

Materials. We used a total of 40 recordings in the experiment. Half of the recordings contained vocals and were performed by popular recording artists, while the other half of the recordings were primarily instrumental (e.g., taken from popular television shows, movies and video games). The mean length of the excerpts was 28.5 s (standard deviation [*SD*] = 5.9 s, range = 13-37 s).[1] A full list of the 40 recordings is shown in Table A1. To select the 40 recordings used in the experiment, we presented 13 undergraduate participants (who did not participate in Experiment 1) with a total of 181 recordings (91 vocal and 90 instrumental). Each participant rated their familiarity with the 181 recordings on a 5-point scale. From these 181 recordings, we selected the 20 highest rated vocal recordings and the 20 highest rated instrumental recordings.

We pitch shifted the recordings using the built-in "Change Pitch" function in Audacity, an open-source program for audio recording and editing. The "Change Pitch" function was developed by Vaughan Johnson and Dominic Mazzoni using "SoundTouch" by Olli Parviainen. The particular pitch-shifting algorithm used by Audacity does not affect overall tempo or speed. Recordings were shifted up or down by one semitone (approximately a 5.9% difference from the original recording) in two 50-cent steps of approximately 2.95% each. Importantly, to minimize the chances of the pitch-shifting process introducing artifacts that could be used in addition to AP cues by participants, we also pitch shifted the in-tune stimuli (either up by 50 cents and then down by 50 cents, or vice versa). Thus, all stimuli presented to participants had undergone the pitch-shifting procedure exactly twice and using the same pitch-shifting magnitude (50 cents). The pitch memory task was coded in E-Prime 2.0 (Psychology Software Tools, Sharpsburg, PA). The INM task was coded in MATLAB (MathWorks, Natick, MA), and the sine tone stimuli used in the INM task were generated in MATLAB and presented to participants at approximately 70 dB SPL. During the experiment, participants listened through Sennheiser HD 280 headphones.

Procedure. After providing informed consent, participants completed the INM task, which has been previously used as a measure of working memory for pitch (Van Hedger et al., 2015). On each trial, participants heard a brief (250 ms) sine wave target note, which was then masked by 1000 ms of white noise. Participants then had to adjust the frequency of a brief (250 ms) starting note (removed from the target note by 1-7 semitones) to try and recreate the originally heard target note. This was achieved by clicking on upward and downward arrows on the computer screen. There were four displayed response arrows: a large upward arrow, a small upward arrow, a large downward arrow and a small downward arrow. Clicking on one of the arrows changed the pitch by either one-third or two-thirds of a semitone upward or downward, depending on whether participants were clicking on the smaller arrows (one-third) or larger arrows (two-thirds). The interim tones, which participants heard while adjusting the starting tone to match the target tone, were also 250 ms in duration. When participants believed that they had successfully recreated the original target note, they pressed a key to move onto the next trial. There were a total of four target notes (F#[4], G[4], G#[4] and A[4]) and eight starting notes (D[4], D#[4], E[4] and F[4] below the target notes and A#[4], B[4], C[5] and C#[5] above the target notes). Target notes could never serve as starting notes on subsequent trials. The entire set of stimuli spanned one octave (excluding the two microtonal steps between the highest starting note, C# and the D from the adjacent octave), meaning there were a total of 34 pitches in the series (including the microtonal traversable notes). Figure 1 provides a sketch of the INM task. Participants randomly heard all combinations of target note/starting notes twice, resulting in 64 trials (4 target notes × 8 starting notes × 2 repetitions).

After completing the INM task, participants completed the main task of judging the pitch of well-known recordings. On each trial, participants heard one of the 40 recordings and then judged whether they thought the song sounded correct or incorrect ("incorrect" was defined to the participants as sounding "too high" or "too low" relative to how they would hear the recording outside of the experiment). Exactly half of the trials contained a correctly tuned recording, while the other half of the trials contained an incorrectly tuned recording (50% of incorrectly tuned recordings were too high by one semitone and 50% were too low by one semitone). Each participant only heard each recording once during the experiment. There were four counterbalanced versions of the experiment, which were presented between participants. After making their pitch judgment on each trial, participants were asked to rate their familiarity with the recording through two questions. The first questions measured whether participants had *any* familiarity with the recording (responding with a forced choice of "yes" or "no"). The second question collected a more graded familiarity response, ranging from 1 to 5. The purpose of these familiarity measures was twofold. First, based on the response from the first question, we discarded trials for which participants reported absolutely no familiarity. Second, based on the second question, we assessed whether greater reports of familiarity with recordings would result in more accurate pitch judgments.

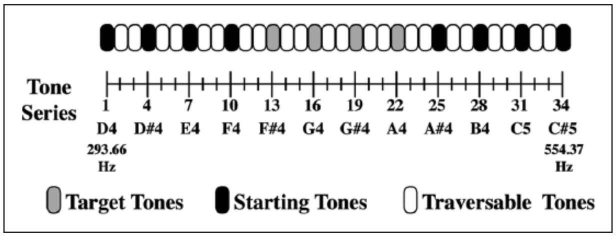

Figure 1. Distribution of target tones and starting tones for the INM task. There were 34 total tones in the distribution, and each tone was separated by one-third of a semitone (approximately 33 cents, or 1.9%). Every target tone/starting tone combination was presented twice, for a total of 64 trials.

After completing the main pitch judgment task, participants filled out several questionnaires, including a music experience questionnaire and the Clarity of Auditory Imagery Scale (CAIS; Willander & Baraldi, 2010). These questionnaires were administered to assess whether explicit musical training or self-reported auditory imagery ability might affect pitch memory for well-known recordings. After completing the questionnaires, participants were debriefed and given course credit.

Results

We first tested whether participants were above chance at judging the pitch of the recordings. We discarded any trials in which participants reported no familiarity with the stimulus. Since the 40 recordings we tested were preselected based on a pilot assessment of popularity among undergraduates, completely unknown recordings comprised relatively few overall trials (3.76 of 40, or 9.40%, participant recognizability range of 50%-100%). Almost every participant (27 of 29) reported knowing over 80% of the presented recordings. Overall, participants made a correct pitch judgment 60.9% of the time, which was above the chance estimate of 50% ($t(28)=6.12, p<0.001$).

Moreover, while we found that the vocal stimuli from recording artists were more accurate than the primarily instrumental theme songs ($t(28)=4.25$, $p<0.001$), both types of recordings were independently above chance (vocal: 66.8%, $t(28)=7.07, p<0.001$; instrumental: 54.5%, $t(28)=2.12$, $p=0.04$). Given these results, we collapsed across vocal and instrumental subsets for all further analyses. These results are plotted in Figure 2a, and the correlations between pitch memory accuracy and the other measured variables are shown in Figure 3a.

Descriptive statistics for the other measured variables (CAIS, INM and music training) are reported in Table 1

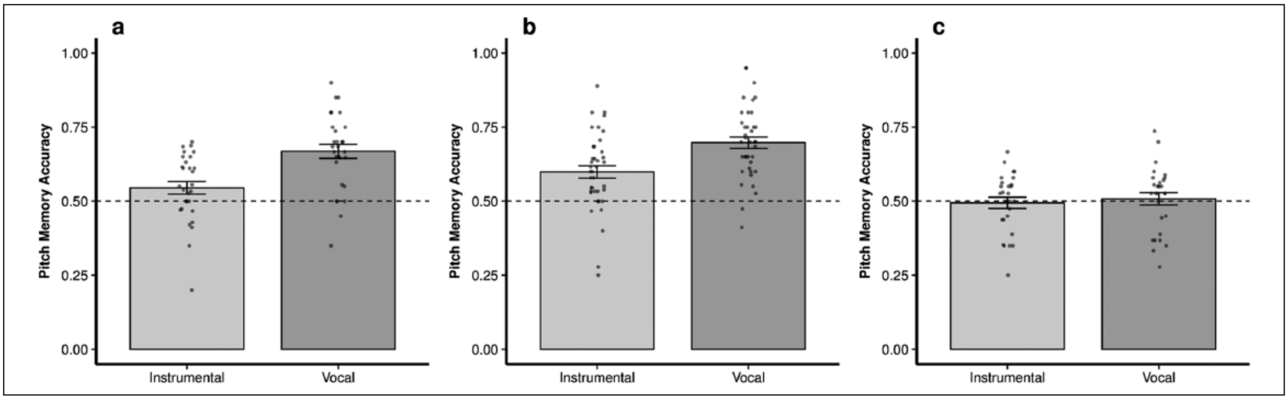

Figure 2. Mean scores for pitch memory accuracy across vocal and instrumental recordings in (a) Experiment 1, (b) Experiment 2 and (c) Experiment 3. Points represent individual participants.

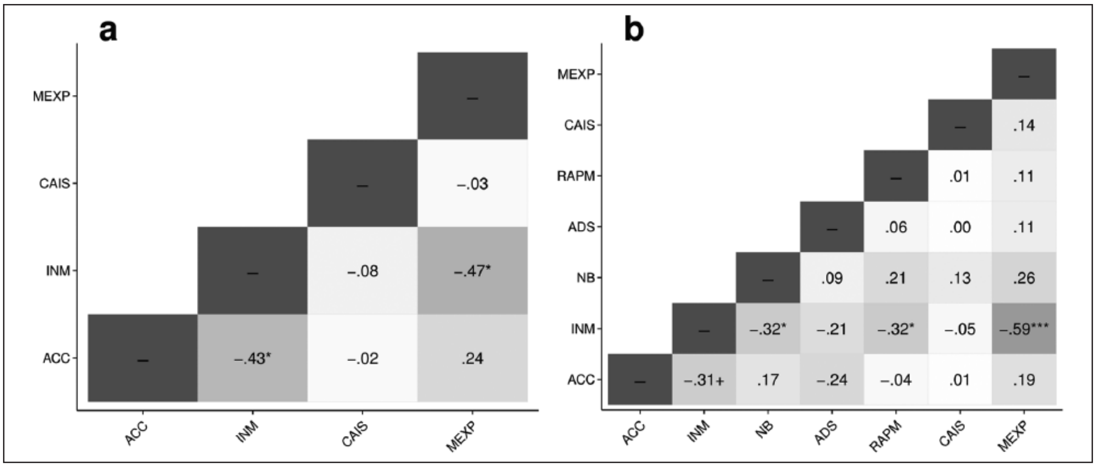

Figure 3. Correlation matrix between pitch memory accuracy (ACC) and all other measured variables for (a) Experiment 1 and (b) Experiment 2. Values in each cell represent Pearson's *r*.
$+0.10 < p \leq 0.05$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 1. Descriptive statistics and fixed effects from the generalized linear mixed-effects model (GLMM) in Experiment.

Descriptive statistics			
Measure	Mean	Standard error	Range
CAIS	3.81	0.14	[2.12, 5]
Music training	6.79	0.83	[0, 15]
INM	3.62	0.27	[0.84, 6.81]
Familiarity	3.57	0.10	[2.26, 4.97]
GLMM fixed effects			
Measure	Estimate	*z*	*Pr*(|z|)
CAIS	−0.035	−0.358	0.350
Music training	0.005	0.310	0.757
INM	−0.134	−2.314	0.021
Familiarity	0.150	2.488	0.013

CAIS: Clarity of Auditory Imagery Scale; INM: implicit note memory. Both INM score and recording familiarity were significant predictors of pitch memory accuracy.

(top). For the CAIS, we used a composite score (averaging across all scale items). For the INM, we took the mean absolute deviation between a participant's final location to which they moved the starting tone in pitch space and the true target note. For example, if a participant's target note was [G4] and her final location in recreating this [G4] was [G#4], she would be three 33-cent steps from the true location and thus receive a score of "3" on that particular trial. Trials in which participants were more than three *SD*s removed from their mean difference score were discarded. These outlier trials comprised very few (1.2%) of the overall trials. We collapsed across all trials, calculating a single INM score per participant. For musical training, we used the maximum reported number of years of active musical instruction on a single instrument. Thus, if a participant had reported 4 years of active piano instruction and 3 years of active violin instruction, they would receive a value of "4."

Predicting individual differences in pitch memory for recordings. To assess how individual differences accounted for variance in long-term pitch memory performance for judging familiar songs, we constructed a generalized linear mixed-effects model (GLMM). The primary benefit of using a mixed-effects model in the present design was that we were able to model both fixed and random effects without having to collapse across our repeated observations (i.e., each presented recording). In particular, this was important for understanding how recording familiarity relates to pitch memory accuracy as more traditional analyses would require collapsing across recording (obtaining an average familiarity rating per participant) or collapsing across participant (obtaining an average familiarity rating per recording). In a mixed-effects model, however, both participant and item variation can be accounted for in the same model.

INM performance, the mean rating from each individual's CAIS questionnaire, the number of years of musical training and familiarity of each recording were fixed effects, while participant and recording stimulus were treated as random effects. In this model, both familiarity and INM performance significantly predicted pitch memory performance. No other fixed effect was significant. The fixed effects of the model are reported in Table 1 (bottom).

Discussion

The present experiment clearly demonstrates that individuals are able to accurately judge when familiar recordings have been shifted in pitch, even when the magnitude of the pitch shift is the smallest conventional pitch difference found in Western music. These findings conceptually replicate previous work (Schellenberg & Trehub, 2003) demonstrating that good pitch memory is widespread in the population, although the present experiment deviates in at least two important ways from previous research. First, we employed an experimental design in which we only presented one version of each recording, not two versions. Thus, participants did not know whether they would ever hear a "correct" version of the recording in our experiment, whereas in previous research participants knew that one of the two versions of each recording they heard would be "correct" on each trial. Second, we used both instrumental and vocal recordings in the present experiment. Despite these deviations, we found similar effect sizes compared to previous research (61% accuracy in the present experiment compared to 58% in Schellenberg & Trehub, 2003).

While we did not find any evidence that explicit musical training or self-reported clarity of auditory imagery explained variance in long-term pitch memory, both recording familiarity and INM performance significantly explained variance in long-term pitch memory. The fact that self-rated recording familiarity explained differences in pitch memory performance is perhaps not surprising, especially since this kind of pitch memory inherently requires at least some familiarity with the to-be-judged stimulus. The relationship between working memory for pitch and long-term AP memory for well-known songs is less obvious, though it has been hinted at in previous work (Jakubowski & Müllensiefen, 2013) albeit as a measure of relative pitch ability. While the INM task employed in the present experiment certainly can be argued to have a relative pitch component (as participants attempt to recreate a target tone from a varying starting tone), based on our results we believe it is more likely that the important factor in explaining pitch memory variance is not relative pitch processing per se, but rather auditory working memory. Indeed, the general procedure of adjusting one note to match a secondary note as closely as possible has been independently used as a measure of pitch precision in auditory working memory (Kumar et al., 2013). Thus, Experiment 1 demonstrates that in addition to recording familiarity, the ability to maintain an accurate pitch representation in working memory is related to one's long-term implicit pitch memory.

Despite the observed relationship between auditory working memory and long-term pitch memory, there are several limitations to Experiment 1, mainly related to the INM task. One prominent limitation of Experiment 1 is that the INM task is not a standard measure of auditory working memory, and thus, its relationship to more canonical measures of auditory working memory is unclear. While some of our prior work has found that the INM task shares a small but significant amount of variance (around 25%) with an auditory n-back task using verbal letter stimuli (Van Hedger et al., 2015), it is presently unclear whether auditory working memory tasks that do not require attending to and manipulating pitch would similarly explain long-term memory for music recordings. It is also possible that the presentation of a backward noise mask and the successive presentation of interim tones to match against the target tone (varying in number depending on the relative distance between the starting tone and target tone) may not have completely masked the sensory properties of the stimulus. As a result, there may be a greater influence of auditory sensory encoding in the INM task than in other working memory tasks. These limitations of the INM task, combined with the fact that no other auditory working memory test was administered to participants in Experiment 1, make it especially difficult to draw strong conclusions about a general relationship between auditory working memory and long-term pitch memory. As such, Experiment 2 was designed to provide a more comprehensive account of the individual differences that relate to long-term pitch memory.

Experiment 2

While the results of Experiment 1 suggest that working memory for pitch may explain individual variance in judging the pitch level of familiar recordings, the scope of the experiment was limited by three important considerations. First, it is unclear whether non-pitched measures of auditory short-term and working memory would similarly explain variance in long-term pitch memory. This is a particularly important consideration, as previous research has found links between forward auditory digit span, auditory n-back and *explicit* AP ability (Deutsch & Dooley, 2013; Van Hedger et al., 2015), although it is currently unclear to what extent these non-musical measures would relate to *implicit* AP ability. Thus, we included a forward auditory digit span and an auditory n-back task in Experiment 2.

A second important limitation of Experiment 1 stems from the typically observed correlation between working memory and fluid intelligence (e.g., Kane, Hambrick, & Conway, 2005). Given the close alignment between working memory and fluid intelligence, our observed relationship between working memory and long-term pitch memory in Experiment 1 may have actually been mediated through fluid intelligence. To assess the how fluid intelligence relates to pitch memory, we have thus included a test of fluid intelligence (the RAPMs) in Experiment 2.

Third, given that the experimental design of Experiment 1 did not alter the ordering of the tasks, it is possible that our results were in part influenced by participants always completing the working memory task before the long-term pitch memory task. While this explanation may seem unlikely, we address this limitation in Experiment 2 through counterbalancing the order of the tasks.

Method

Participants. In total, 40 University of Chicago undergraduates participated in the study for course credit ($M_{\text{age}}=$ 20.03 years, $SD_{\text{age}}=1.12$ years, range = 18-22 years, 28 females, 12 males). Participants were not specifically recruited for their musical experience, though a majority of participants (28 of 40) reported at least some music instruction. Further details of participants' musical training are reported in section "Results." One participant reported possessing AP, leaving 39 analyzable participants.

Materials. The 40 recordings used in the experiment were the same as those used in Experiment 1. As such, the pitch-shifting procedure was identical to Experiment 1. The pitch memory task, the auditory n-back task and the auditory digit span task were all coded in E-Prime 2.0 (Psychology Software Tools). The RAPM was administered in a timed, paper–pencil format. The INM task was coded in MATLAB (MathWorks), and the sine tone stimuli used in the INM task were generated in MATLAB and presented to participants at approximately 70 dB SPL. During the experiment, participants listened through Sennheiser HD 280 headphones.

Procedure. Each experimental session began with either the battery of cognitive tests (the INM, auditory n-back, auditory digit span and the RAPM, in a randomized order) or the pitch memory task. Half of the participants completed the cognitive battery first, while half of the participants completed the pitch memory task first. After both the cognitive battery and the pitch memory components had been completed, participants filled out the CAIS and musical experience questionnaires. The INM task, the pitch memory task and the two questionnaires (CAIS and musical experience) were administered using the same procedures as Experiment 1.

For the auditory digit span task, participants completed a total of 18 trials. The digit span was non-adaptive, in that the number of digits increased throughout the task regardless of performance. The initial digit level was set at 5 and the final digit level set at 12. Participants completed two trials at each digit level, with the exception of the initial digit level in which participants completed four trials. Digits were presented at a rate of one every 2 s. We operationalized digit span performance as the number of total correct trials per participant.

The auditory n-back required participants to monitor spoken letter strings and respond when the current spoken letter was the same as the one presented "*n*" trials previously. All participants completed a short practice, as well as a main task, in which "*n*" was fixed at 3. Participants responded to all spoken letters (i.e., pressing a button labeled "target" or "not target" for each letter). The task consisted of three runs of 30 letter strings (10 targets and 20 non-targets per run, resulting in 30 total targets and 60 total non-targets). Letters were spoken every 3000 ms. We calculated d' scores for each participant. A perfect score (30 of 30 hits and 0 of 60 false alarms) would result in a d' score of 4.52. Three participants achieved perfect scores.

The RAPM test was administered in a booklet. The test was introduced as a series of pattern completion problems. After completing two practice items from Set I in an untimed format, participants completed as many items from Set II as possible in 20 min (36 total items). While this time limit is aggressive, it was based on prior work in which participants were given 10 min to complete half of Set II (Jaeggi et al., 2010). Participants were told when they had 15, 10 and 5 min remaining. No participant completed all items in the test.

Results

Similar to Experiment 1, we discarded any trials in which participants reported no familiarity with the recording. Unknown recordings comprised relatively few overall trials (4.97 of 40, or 12.43%, participant recognizability range of 62.5%-100%). Well over half of the participants (29 of 39) reported knowing over 80% of the presented recordings. This recognition rate was slightly lower than what was observed in Experiment 1, though it should be noted that the pilot test for recording familiarity was more closely aligned in time with the participants from Experiment 1 compared to Experiment 2. Despite this slight decrease in recording recognition, we replicated the findings from Experiment 1, as participants made correct pitch judgments 65.1% of the time, which was above the chance estimate of 50% ($t(38)=10.61$, $p<0.001$).

While we also replicated the finding from Experiment 1 that the vocal stimuli from recording artists were more accurate than the primarily instrumental theme songs ($t(38)=3.39$, $p=0.002$), both types of recordings were independently above chance (vocal: 69.8%, $t(38)=10.23$, $p<0.001$; instrumental: 59.9%, $t(38)=4.66$, $p<0.001$). Given these results, we collapsed across vocal and instrumental subsets for all further analyses. These results are plotted in Figure 2b, and the correlations between pitch memory accuracy and the other measured variables are displayed in Figure 3b.

The mean, standard error and range of participant responses for all other collected measures—CAIS (composite score), music training (years), INM (number of steps removed from target note), auditory digit span (number of correct trials), auditory n-back (d') and RAPM (number of correct items) are reported in Table 2 (top).

Predicting individual differences in pitch memory for recordings. To assess how individual differences accounted for variance in long-term pitch memory performance for judging familiar recordings, similar to Experiment 1, we

Table 2. Descriptive statistics and fixed effects from the generalized linear mixed-effects model (GLMM) of Experiment 2.

Descriptive statistics

Measure	Mean	Standard error	Range
CAIS	3.95	0.09	[2.88, 5]
Music training	5.40	0.81	[0, 15]
INM	3.93	0.27	[1.23, 8.24]
Digit span	9.44	0.34	[5, 15]
N-back	3.26	0.13	[1.37, 4.52]
Raven's	23.7	0.56	[17, 31]
Familiarity	3.47	0.08	[2.50, 4.36]

GLMM fixed effects

Measure	Estimate	z	*Pr*(|z|)
CAIS	−0.024	−0.209	0.835
Music training	−0.002	−0.135	0.892
INM	−0.148	−2.729	0.006
Digit span	−0.035	−1.056	0.291
N-back	0.073	0.802	0.423
Raven's	−0.027	−1.296	0.195
Familiarity	0.363	6.133	<0.001

Similar to Experiment 1, we found that both INM score and recording familiarity were significant predictors of long-term pitch memory.

constructed a GLMM. We included seven fixed effects: (1) INM performance, (2) auditory n-back performance, (3) auditory digit span, (4) RAPM score, (5) the mean rating from each individual's CAIS questionnaire, (6) the number of years of musical training and (7) the familiarity of each recording. Participant and recording stimulus were treated as random effects. In this model, recording familiarity and INM score were significant predictors of pitch memory. The fixed effects of the model are reported in Table 2 (bottom).

Discussion

Experiment 2 followed up on some alternative explanations of pitch memory performance from Experiment 1—specifically, (1) whether the relationship between auditory working memory and long-term pitch memory could be explained by more general cognitive factors, (2) whether non-pitched measures of auditory working memory and short-term memory could similarly explain variance in long-term pitch memory and (3) whether task ordering influences the relationship between auditory working memory and long-term pitch memory. To specifically address these questions, we included a measure of fluid intelligence (the RAPMs), we included two additional measures of auditory working memory and short-term memory that were not musical in nature (the auditory n-back with spoken letters and the auditory digit span), and we randomized whether participants completed the cognitive measures versus the long-term pitch memory task first.

Experiment 2 replicates the positive relationship between auditory working memory and long-term pitch memory found in Experiment 1, while also providing some important clarifications regarding this relationship. First, it does not appear that the relationship between auditory working memory and long-term memory for pitch can be best explained by general (non-auditory) cognitive ability, as performance on the RAPMs was not significantly related to long-term pitch memory, and the relationship between INM performance and long-term pitch memory remained significant when controlling for RAPM performance. Second, the relationship between auditory working memory and long-term pitch memory may be limited to auditory working memory for pitch, as performance on the auditory n-back—which used spoken letters as stimuli—was not significantly related to long-term pitch memory. Third, the relationship between working memory for pitch and long-term pitch memory did not appear to be an artifact of task presentation as half of the participants in the present experiment completed the long-term pitch memory task first, while the other half completed the battery of cognitive tests first, and we still found evidence that INM performance was significantly explaining long-term pitch memory variance.

Experiment 3

To confirm that the results from Experiments 1 and 2 could not be attributed to confounding factors as a result of the pitch-shifting procedure, Experiment 3 adopted the same general structure as Experiments 1 and 2, although the recording we used were preselected for their *unfamiliarity*. If the above-chance performance for judging the pitch of recordings in Experiments 1 and 2 was truly driven by recording familiarity, then participants should be at chance at selecting the correct tuning of unfamiliar recordings. If, however, we had introduced unwanted cues as a result of the pitch-shifting procedure, then participants might still be above chance. While this possibility seems unlikely, especially since *all* songs in Experiment 1 underwent the pitch-shifting procedure, it is possible that pitch shifting specifically vocal stimuli resulted in significant timbral changes (e.g., changing the voice characteristics of the singer) that could have been used in addition to long-term pitch memory. It also should be noted that this general approach (presenting unfamiliar pitch-shifted recordings to confirm pitch memory is not confounded by the pitch-shifting procedure) has been previously used in studies of long-term pitch memory (Schellenberg & Trehub, 2003).

Method

Participants. In total, 28 University of Chicago undergraduates participated in the study for course credit (M_{age}= 19.79 years, SD_{age} = 1.13 years, range = 18-22 years, 20 females, 8 males). Participants were not specifically recruited for their musical experience, although a majority of participants (21 of 28) reported at least some music instruction. The mean number of years participants reported playing their primary instrument was 6.01 (*SD*=4.96 years, range=0-14 years).

Materials. The 40 recordings used in the experiment were selected by the experimenters based on the following criteria. First, we selected 20 vocal stimuli from recording artists and 20 instrumental themes similar to Experiments 1 and 2. Second, the vocal recordings we selected were from popular recording artists, even though the songs themselves were selected by the first authors based on their presumed obscurity. This presumed obscurity was objectively assessed through examining the number of listens each recording had accrued on Spotify (if it existed on Spotify), which had to be fewer than 1 million at the time of the experiment to be considered. The reason we selected well-known recording artists, including some artists who were used in Experiment 1, is that we wanted to assess whether the pitch-shifting process would significantly alter the timbre of artists' voices, thereby allowing participants to make accurate pitch judgments despite not recognizing the recording. A full list of the 40 songs is printed in Table A2. The mean length of the excerpts was 29.4 s (*SD*=4.0 s, range=11-39 s).[2] We pitch shifted the recordings in an identical manner as in Experiments 1 and 2. The experiment was coded in E-Prime 2.0 (Psychology Software Tools). During the experiment, participants listened through Sennheiser HD 280 headphones at approximately 70 dB SPL.

Procedure. Participants only completed the pitch memory task. We did not include any additional measures or questionnaires, as we hypothesized that participants should be at chance at judging the pitch of unfamiliar recordings. As such, we did not anticipate being able to explain individual differences in pitch memory accuracy.

Results

We discarded any trials in which participants reported having any familiarity with the recording. Since the 40 recordings were selected based on presumed obscurity among undergraduates, recognized recordings comprised relatively few overall trials (2.39 of 40, or 5.98%, participant range of 0%-22.5%). Almost every participant (27 of 28) reported knowing fewer than 20% of recordings. As hypothesized, we did not find any evidence that participants were able to correctly determine the pitch of unfamiliar recordings. Overall, participants made a correct pitch judgment 50.2% of the time, which was not above the chance estimate of 50% ($t(27)=0.13$, $p=0.896$). Figure 2c displays the mean accuracy and individual participant observations for both vocal and instrumental recordings. We did not find any evidence that vocal recordings were more accurate than the instrumental recordings ($t(27)=0.61$, $p=0.548$), which is not surprising as both kinds of recordings were not independently above the chance estimate of 50% (vocal: 50.8%, $t(27)=0.40$, $p=0.693$; instrumental: 49.5%, $t(27)=-0.29$, $p=0.77$).

Given that the previous analyses rest on accepting null results, we assessed whether performance in the present experiment was significantly different from performance in Experiments 1 and 2. Independent samples *t*-tests showed that overall performance ($t(94)=6.34$, $p<0.001$), vocal recording performance ($t(94)=6.54$, $p<0.001$) and instrumental recording performance ($t(94)=3.02$, $p=0.003$) were significantly worse in the present experiment compared to Experiments 1 and 2.

Discussion

Experiment 3 assessed whether particular aspects of the experimental design of Experiments 1 and 2 introduced artificially high performance. First, we wanted to rule out the possibility that the pitch-shifting procedure introduced unwanted auditory cues that could be used in conjunction with pitch memory. Second, given that vocal stimuli have only very recently been used to assess long-term pitch memory (Jakubowski, Müllensiefen, & Stewart, 2017), we wanted to assess whether the results of Experiments 1 and 2 were in part driven by timbral-related changes in pitch-shifting familiar vocalists.

The results from Experiment 3 demonstrate that hearing recordings in consistent musical keys is crucial for instantiating long-term pitch memory. Participants could not determine the correct pitch height of unfamiliar vocal and instrumental songs at above-chance levels. This general finding replicates prior research arguing that at least some stimulus familiarity is essential for accurate pitch memory (e.g., Schellenberg & Trehub, 2003). Experiment 3 also provides a conceptual replication of recent work demonstrating that vocal recordings can be used to assess long-term pitch memory (Jakubowski et al., 2017), although the results from the present experiment go beyond this prior research by demonstrating that vocal recordings *by well-known artists* similarly cannot be identified if the recording is not known. In this sense, Experiment 3 provides a methodological advance in the study of pitch memory for familiar musical recordings, clearly demonstrating that at least within a one-semitone pitch-shifting range (approximately a 5.9% change) unfamiliar vocal recordings of known singers can be used to assess pitch memory.

General discussion

How is it that we can reliably judge when we are hearing a well-known recording at its correct pitch height,

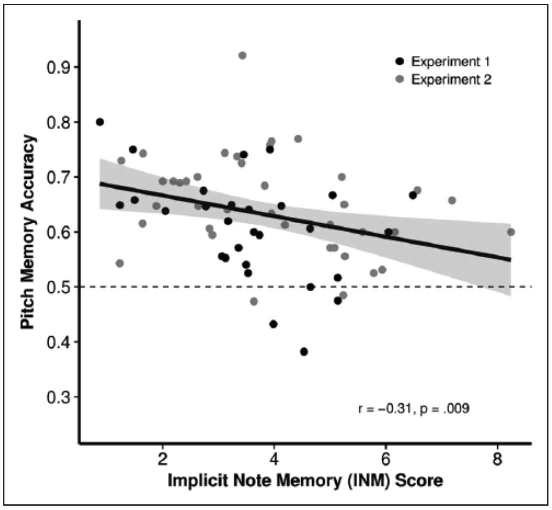

Figure 4. Overall relationship between INM score and pitch memory for recordings in Experiments 1 and 2.

independent of explicit AP labels? The present set of experiments shed light on the mechanisms underlying the formation of these pitch representations. Specifically, while the present set of experiments clearly demonstrates that familiarity is important in the formation of accurate long-term pitch representations (shown both through the relationship between familiarity ratings and performance in Experiments 1 and 2, as well as the at-chance performance in Experiment 3), the experiments suggest the presence of another important factor in pitch memory accuracy—auditory working memory for pitch. Figure 4 displays the overall relationship between the INM task and pitch memory across Experiments 1 and 2 ($n=68$).

Why might the short-term ability to adjust a starting tone to match a target tone relate to the ability to judge when a popular recording is presented in its most commonly heard musical key? The answer to this question is perhaps best answered by reassessing the particular construction of working memory. While the broad definition of working memory is the ability to temporarily hold onto information in a mental workspace—a process requiring attention (Fougnie, 2008)—recent research has begun to dissociate the precision with which items can be maintained in working memory from the capacity or number of items able to be maintained in working memory (e.g., Zhang & Luck, 2011). This view of working memory is distinct from flexible-resource theories of working memory, which propose no functional distinction between precision and capacity as individuals could hold either a small number of items in working memory with high precision or a large number of items in working memory with poor precision (Bays & Husain, 2008; Palmer, 1990; Wilken & Ma, 2004). In vision, separate neural mechanisms have been reported with regard to working memory capacity and precision (Xu & Chun, 2006).

To the extent that auditory precision and capacity can be thought of as separable aspects of working memory (e.g., see Ku, Bodner, & Zhou, 2015), the present results support the notion that working memory *precision* may be the best working memory construct in explaining variance in long-term pitch memory. The INM task is putatively a stronger measure of auditory working memory precision, as it involves the precise recreation of a single auditory item after a delay period. Moreover, the task of judging the pitch height of well-known recordings is similarly centered on auditory precision, albeit in a long-term memory store. Thus, the present results suggest that better auditory working memory precision for musical pitch potentially allows individuals to learn and remember the AP height of well-known recordings, even if this learning is incidental (i.e., results from passive listening to recordings with no instruction to remember pitch height). To further support the dissociation of working memory precision and capacity in explaining variance in long-term pitch memory, we did not find evidence that the auditory n-back or auditory digit span from Experiment 2—putative measures of capacity—was significantly related to long-term pitch memory. Yet, from the current results, it is unclear whether this was solely because the auditory n-back and digit span are capacity tasks or whether it was partly because of the nature of the stimuli used (specifically, spoken letters and numbers over musical notes). Future work could provide a more comprehensive account of this relationship through manipulating the constituent auditory tokens of each task (e.g., using pitches in the INM that do not correspond to Western musical notes, using speech sounds in the INM instead of musical notes or using musical notes in either the auditory n-back or digit span). These kinds of manipulations would further our understanding of the extent to which these findings are due to dissociations of precision and capacity in working memory.

Another unresolved question from the present experiments has to do with the particular nature of the INM task, in which both sensory and working memory could potentially contribute to performance. As such, it is unclear whether individual differences in sensory memory may additionally explain individual variance in long-term pitch memory. Recent research has demonstrated that individual differences in low-level, sensory performance are related to higher level abilities (Albouy, Cousineau, Caclin, Tillmann, & Peretz, 2016). Moreover, the INM task can be conceptualized as containing both a sensory memory component (initial comparison of target tone to starting tone) and a working memory component (active recreation of the target tone while cycling through interim tones). As such, future research exploring individual differences in long-term pitch memory may benefit from attempts to disentangle sensory memory from working memory.

While the relationship between self-reported recording familiarity and pitch memory accuracy found in Experiments

1 and 2 seems obvious, it is interesting to note that previous research on pitch memory for recordings have generally failed to find a strong graded relationship between familiarity and accuracy (e.g., Jakubowski et al., 2017; Schellenberg & Trehub, 2003). There are two likely reasons why we found such a strong relationship between familiarity and pitch memory accuracy in the present experiments. First, the nature of our paradigm included a much larger distribution of recordings, which means that we were likely able to capture a greater range of familiarity estimates among recordings that were, nevertheless, recognizable. Second, by framing the question in terms of familiarity, rather than something more objective (such as estimated number of viewed episodes for television themes, used in Schellenberg & Trehub, 2003), it is possible that participants were using familiarity ratings—which were always presented after making a pitch height judgment—as a post hoc confidence rating.

How can these results be interpreted in the larger framework of AP memory formation for both AP and non-AP listeners? On one hand, the explicit knowledge of pitch chroma labels (genuine AP) appears to be an extremely rare ability, putatively occurring in less than one in every 10,000 individuals (Bachem, 1955). Genuine AP is described as dichotomous, with clear and easy-to-define “AP possessors” and “non-AP possessors” (Athos et al., 2007; though see Bermudez & Zatorre, 2009). On the other hand, accurate pitch memory for familiar sounds, including but not limited to musical recordings, appears widespread and normally distributed in the population (e.g., Schellenberg & Trehub, 2003; Van Hedger et al., 2016). Given these stark differences, it might appear that there is no real relationship between these abilities, even if both are concerned with memory for AP. If, however, genuine AP can be conceptualized as a two-step process, with general pitch memory ability constituting the first step and explicit categorization based on pitch chroma constituting the second step (Levitin & Rogers, 2005), the discussion of implicit pitch memory in service of genuine AP may prove to be important in better understanding both abilities. For example, genuine AP possessors appear to have better pitch memories for well-known stimuli (Dooley, 2011), even when the use of explicit AP labels is not beneficial to the task (Van Hedger et al., 2016), suggesting that general long-term pitch memory among AP possessors may be enhanced relative to non-AP possessors. Moreover, features of the INM task (such as holding onto a single pitch in memory while listening to intermediary pitches) have been previously used to demonstrate differences between AP and non-AP possessors (Siegel, 1974) and even to claim that AP can exist independently of explicit musical note labels (Ross, Olson, & Gore, 2003). Thus, the present results may help bridge the literature on both explicit and implicit AP abilities.

Taking into account individual differences—particularly in auditory working memory—might help explain the relationship between pitch memory among both AP and non-AP possessors. Recent work has found that individual differences in auditory working memory precision, using the same INM task that was used in Experiment 1, explained individual differences in the *explicit* acquisition of AP categories (Van Hedger et al., 2015). Combined with the results of the present experiments, this suggests that auditory working memory precision might constitute a meaningful individual difference for explaining the ability to represent AP information in long-term memory, whether that information is represented implicitly (e.g., having a general feeling that a familiar recording sounds “off”) or explicitly (e.g., learning to call a frequency of 440 Hz an “A”). However, unlike the current set of experiments, it is important to note that auditory n-back was a significant predictor of explicit AP learning in Van Hedger et al. (2015). Why would both INM and auditory n-back explain variance in explicit AP category learning, while only INM explained variance in implicit pitch memory in the present experiments? There are at least two reasons. First, the explicit acquisition of 12 note categories may place additional demands on working memory (particularly on capacity) in a way that pitch judgments of familiar recordings do not. Second, as alluded to earlier, the relatively strong correlations observed in the present experiments between INM performance and explicit musical training make it difficult to disentangle working memory precision from long-term note category structures that may be established through years of musical training. More specifically, it could be the case that musical training improves implicit note representations (cf. Heald, Van Hedger, & Nusbaum, 2014; Van Hedger et al., 2017) which could then influence performance on the INM task. This kind of explanation would still be interesting as it would suggest that the AP memory traces resulting from primarily incidental exposure to popular recordings could be influenced by the preexisting strength of implicit note categories. However, this kind of explanation would also require a reconsideration of whether auditory working memory precision—broadly conceived—can be thought to relate to long-term pitch memory.

In summary, the present set of experiments conceptually replicates previous work demonstrating that pitch memory is widespread and normally distributed in the general population (e.g., Schellenberg & Trehub, 2003). However, the present work extends these findings by demonstrating that individual differences in auditory working memory for pitch are related to long-term pitch memories of well-known recordings. This relationship between auditory working memory and long-term memory fidelity can be potentially interpreted in the context of recent empirical and theoretical work in the visual domain (Brady, Konkle, Gill, Oliva, & Alvarez, 2013), suggesting that pitch fidelity in auditory working memory and long-term memory may also be bounded by a common limit between both kinds of memory. Finally, the recent distinction between working memory precision and capacity allows for a more nuanced understanding of how auditory working memory performance might differentially matter in the formation of implicit and explicit AP representation. Furthering our understanding how both auditory working memory capacity and precision influence implicit and explicit long-term pitch representations will undoubtedly help us understand the general mechanisms underlying pitch memory across populations.

Acknowledgements

The authors would like to thank Luke Clohisy for his assistance in data collection.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Multidisciplinary University Research Initiatives (MURI) Program of the Office of Naval Research grant (grant number: DOD/ONR N00014-13-1-0205).

Supplementary material

Supplementary Tables A1 and A2 are available at http://journals.sagepub.com/doi/suppl/10.1080/17470218.2017.1307427.

Notes

1. All 20 of the vocal excerpts from recording artists were 30 s in duration and were chosen by undergraduate research assistants to represent the most iconic portion of the recording. This overwhelmingly resulted in the selection of the chorus (the first time it occurs), with approximately two measures of antecedent music. The only exceptions to this rule were "Hey Jude" and "Bohemian Rhapsody," in which the opening of the song was determined by undergraduate research assistants to be the most iconic. The 20 television, movie and video game themes varied based on the overall length of the theme. For example, the theme from the television show "30 Rock" was only 17 s in duration, while the theme from "Friends" (I'll Be There for You) was 36 s in duration. For themes that were significantly longer than 30 s (e.g., movie themes), we selected a musical phrase that was as close to 30 s as possible while allowing harmonic resolution or musical phrase completion.
2. Similar to Experiment 1, all 20 of the vocal songs were 30 s in duration. The 20 television, movie and video game themes varied based on the overall length of the theme. For example, the theme from the television show "Ugly Betty" was only 11 s in duration, while the theme from "Teen Wolf" was 39 s in duration. For themes that were significantly longer than 30 s, we selected a musical phrase that was as close to 30 s as possible while allowing harmonic resolution or musical phrase completion.

ORCID iD

Howard C Nusbaum http://orcid.org/0000-0001-8032-7297

References

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Scientific Reports*, *6*, Article 18861. doi:10.1038/srep18861

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 14795–14800. doi:10.1073/pnas.0703868104

Bachem, A. (1955). Absolute pitch. *The Journal of the Acoustical Society of America*, *27*, 1180–1185.

Bartlette, C., Henry, M. L., & Moore, J. (2015). Interaction of relative pitch memory and latent absolute pitch for songs in an ordered list. *Psychomusicology*, *24*, 279–290.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854. doi:10.1126/science.1158023

Beilock, S. L., & Holt, L. E. (2007). Embodied preference judgments: Can likeability be driven by the motor system? Research article. *Psychological Science*, *18*, 51–57. doi:10.1111/j.1467-9280.2007.01848.x

Ben-Haim, M. S., Eitan, Z., & Chajut, E. (2014). Pitch memory and exposure effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 24–32. doi:10.1037/a0033583

Bermudez, P., & Zatorre, R. J. (2009). A distribution of absolute pitch ability as revealed by computerized testing. *Music Perception*, *27*(2), 89–101. doi:10.1525/rep.2008.104.1.92

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*, 981–990. doi:10.1177/0956797612465439

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, *396*(6707), 128. doi:10.1038/24075

Creel, S. C., & Tumlin, M. A. (2012). Online recognition of music is influenced by relative and absolute pitch information. *Cognitive Science*, *36*, 224–260. doi:10.1111/j.1551-6709.2011.01206.x

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*, 1859–1861. doi:10.1121/1.4754917

Dooley, K. (2011). *Absolute pitch and related abilities*. San Diego: University of California, San Diego.

Fougnie, D. (2008). The relationship between attention and working memory. In N. B. Johansen (Ed.), *New research on short-term memory* (pp. 1–45). New York, NY: Nova Science Publishers, Inc. doi:10.3389/conf.fnhum.2011.207.00576

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Mullensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae*, *17*, 334–349. doi:10.1177/1029864913490633

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

Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2014). Auditory category knowledge in experts and novices. *Frontiers in Neuroscience*, *8*, 1–15. doi:10.3389/fnins.2014.00260

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*, 439–450. doi:10.1037/0894-4105.17.3.439

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y. F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—Implications for training and transfer. *Intelligence*, *38*, 625–635. doi:10.1016/j.intell.2010.09.001

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. doi:10.1080/17470218.2013.803136

Jakubowski, K., Müllensiefen, D., & Stewart, L. (2017). A developmental study of latent absolute pitch memory. *The Quarterly Journal of Experimental Psychology*, *70*, 434–443. doi:10.1080/17470218.2015.1131726

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71. doi:10.1037/0033-2909.131.1.66

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*, *12*, 125–143.

Ku, Y., Bodner, M., & Zhou, Y.-D. (2015). Prefrontal cortex and sensory cortices during working memory: Quantity and quality. *Neuroscience Bulletin*, *31*, 175–182. doi:10.1007/s12264-014-1503-7

Kumar, S., Joseph, S., Pearson, B., Teki, S., Fox, Z. V., Griffiths, T. D., & Husain, M. (2013). Resource allocation and prioritization in auditory working memory. *Cognitive Neuroscience*, *4*, 1–9. doi:10.1080/17588928.2012.716416

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

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*, 26–33. doi:10.1016/j.tics.2004.11.007

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE*, *5*(6), e11120. doi:10.1371/journal.pone.0011120

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 332–350. doi:10.1037/0096-1523.16.2.332

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*, 653–661. doi:10.1097/AUD.0b013e3181b412e9

Rakowski, A., & Morawska-Bungeler, M. (1987). In search for the criteria of absolute pitch. *Archives of Acoustics*, *12*, 75–87.

Ross, D. A., Olson, I. R., & Gore, J. C. (2003). Absolute pitch does not depend on early musical training. *Annals of the New York Academy of Sciences*, *999*, 522–526. doi:10.1196/annals.1284.065

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition*, *43*, 1021–1031. doi:10.3758/s13421-015-0519-1

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

Schellenberg, E. G., Stalinski, S. M., & Marks, B. M. (2014). Memory for surface features of unfamiliar melodies: Independent effects of changes in pitch and tempo. *Psychological Research*, *78*, 84–95. doi:10.1007/s00426-013-0483-y

Siegel, J. A. (1974). Sensory and verbal coding strategies in subjects with absolute pitch. *Journal of Experimental Psychology*, *103*, 37–44.

Smith, N. A., & Schmuckler, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. *The Journal of the Acoustical Society of America*, *123*(4), EL77-EL84. doi:10.1121/1.2896106

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*, 345–361. doi:10.1037/0033-2909.113.2.345

Van Hedger, S. C., Heald, S. L. M., Huang, A., Rutstein, B., & Nusbaum, H. C. (2017). Telling in-tune from out-of-tune: Widespread evidence for implicit absolute intonation. *Psychonomic Bulletin & Review*, *24*, 481–488. doi:10.3758/s13423-016-1099-1

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. doi:10.1016/j.cognition.2015.03.012

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2016). What the [bleep]? Enhanced absolute pitch memory for a 1000 Hz sine tone. *Cognition*, *154*, 139–150. doi:10.1016/j.cognition.2016.06.001

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 1120–1135. doi:10.1167/4.12.11

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

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. doi:10.1038/nature04262

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*, 1434–1441. doi:10.1177/0956797611417006