Memory & Cognition
https://doi.org/10.3758/s13421-025-01686-0

Short-term pitch memory predicts both incidentally and intentionally acquired absolute pitch categories

Stephen C. Van Hedger[1,2] · **Katarina Jovanovic**[1] · **Andrè Grenier**[1] · **Sum Yee Hoh**[1]

Accepted: 31 December 2024

Abstract

Tonal short-term memory has been positively associated with both incidentally acquired absolute pitch memory (e.g., for popular songs) and explicitly learned absolute pitch (AP) categories; however, the relationship between these constructs has not been directly tested within the same individuals. The current study investigated how tonal short-term memory relates to both incidentally and intentionally acquired AP. Participants ($n = 192$) completed a tonal short-term memory task, an incidental AP task, and an AP categorization task. The tonal short-term assessment involved adjusting a starting tone to match a target tone. The incidental AP task involved judging whether excerpts of popular songs were presented in the correct key. The AP categorization task involved associating six pitch chroma categories with arbitrary labels, including a generalization test that used Shepard tones to discourage pitch height cues. We found that all three pitch measures were positively correlated with one another. Critically, however, we found that tonal short-term memory fully mediated the relationship between incidental AP and explicit AP categorization. This finding held even when controlling for musical training and tonal language fluency. Overall, these results suggest that pitch memory is a consistent individual difference measure across different timescales and different measures (e.g., incidental measures, explicit measures). However, tonal short-term memory appears to be foundational to both incidentally acquired and explicitly learned AP categories.

Keywords Absolute pitch · Memory · Categorization · Generalization

Absolute pitch (AP) is the ability to identify the name of a musical tone or to produce a musical tone without the use of an external reference (Takeuchi & Hulse, 1993). The acquisition of AP is thought to be constrained by a critical period of development (e.g., see Gervain et al., 2013; Levitin & Rogers, 2005; Levitin & Zatorre, 2003), with some studies reporting that children outperform adults in AP learning paradigms (Crozier, 1997; Russo et al., 2003) and one cross-sectional study finding evidence consistent with widespread trainability of AP among 4-to-10-year-old music students (Miyazaki & Ogawa, 2006). Although AP has been considered to be essentially impossible to train in adulthood due to its presumed dependence on an early critical period of development, recent studies have suggested that most adults who receive explicit AP training show improvements in category labeling, with a small subset of individuals achieving "genuine" levels of AP performance in terms of response speed and categorization accuracy (Van Hedger et al., 2019; Wong, Lui et al., 2020a, Wong, Ngan, et al., 2020b). This large degree of variability in adult AP training outcomes, despite comparable amounts of training, suggests that there are meaningful individual differences in explaining AP category learning, which are of both theoretical and practical importance.

To characterize this variability in AP performance, it is important to consider the underlying components of AP. Specifically, AP is thought to consist of two underlying constructs: (1) *pitch memory*, a type of long-term memory for pitch, incidentally acquired through repeated exposure (e.g., remembering the absolute pitch level of a popular song), and (2) *pitch labeling*, which is the ability to explicitly associate musical notes with labels (e.g., C, or *do* in the Western scale) and is considered the hallmark of AP (Levitin & Rogers,

✉ Stephen C. Van Hedger
svanhedg@uwo.ca

[1] Department of Psychology, Huron University College at Western, 1349 Western Road, London, ON N6G 1H3, Canada

[2] Department of Psychology and Western Institute for Neuroscience, Western University, London, ON, N6A 3K7, Canada

2005). Whereas pitch memory is present among most individuals and is distributed normally among the general population (e.g., Frieler et al., 2013; Levitin, 1994; Lockhead & Byrd, 1981; Schellenberg & Trehub, 2003), pitch labeling is considerably rarer, although there is debate as to whether it is distributed dichotomously (Athos et al., 2007) or more continuously (Bermudez & Zatorre, 2009; Van Hedger et al., 2020) in the population.

Given this two-component model of AP, the identification of relevant individual differences related to AP performance might depend on whether one is assessing incidentally acquired *pitch memory* or explicitly acquired *pitch labels*. Despite the presumed association between pitch memory and pitch labeling (Levitin & Rogers, 2005), previous research has suggested some degree of dissociation in terms of the relevant individual difference measures between the two constructs. For pitch labeling, there have been several identified factors that have been associated with the development and maintenance of AP (e.g., see Wilson et al., 2012). Some of the more consistent factors associated with successful AP acquisition are early musical training (e.g., Levitin & Zatorre, 2003; Russo et al., 2003), speaking a tonal language (Deutsch et al., 2004, 2006), and Asian ethnicity (Gregersen et al., 1999). Yet, individual variability in incidental pitch memory (e.g., for popular songs) has not been found to relate to musical training (Jakubowski & Müllensiefen, 2013; Schellenberg & Trehub, 2003; Van Hedger et al., 2018, 2024), speaking a tonal language, or Asian ethnicity (Schellenberg & Trehub, 2008). Incidental pitch memory was also not found to be related to a pitch decay task that is thought to be a valid proxy for pitch labelling ability (Jakubowski & Müllensiefen, 2013), further suggesting that these abilities might be dissociable. Although one potential explanation of this dissociation is that pitch memory and pitch labeling might be tapping into different memory representations for pitch (i.e., a representation based on pitch chroma for labeling and a representation based on pitch height for memory), recent work has shown that pitch memory for popular songs generalizes across timbre and octave (Van Hedger et al., 2023), which suggests that both abilities are grounded in representations based on pitch chroma.

One promising construct that has been associated with both explicitly acquired AP categories and incidental pitch memory is auditory short-term memory. Within the domain of pitch labeling, there are several complementary pieces of evidence that support a role of auditory short-term memory in forming pitch labels. First, AP possessors have been shown to outperform musically matched, non-AP controls on auditory (but not visual) digit span tests irrespective of age (Deutsch & Dooley, 2013). This finding suggests that AP possessors have significantly enhanced auditory short-term memory relative to non-possessors, with the authors speculating that this difference in auditory short-term memory might partly underlie AP. Second, when AP possessors are given more challenging assessments (e.g., using more extreme octave ranges and difficult-to-identify timbres), performance variability relates in part to auditory working memory, assessed via an *N*-back using spoken letters (Van Hedger & Nusbaum, 2018). Third, the ability of non-AP possessors to explicitly acquire pitch labels based on chroma has been associated with auditory short-term memory in both rapid (Van Hedger et al., 2015) and more extensive (Van Hedger et al., 2019; Wong, Ngan et al., 2020b) training paradigms, with the more extensive paradigms producing a small subset of individuals who exceeded typical thresholds for determining "genuine" AP ability. Taken together, these results suggest that auditory short-term memory is important in the development of explicit labels based on pitch chroma.

Auditory short-term memory has also been positively associated with individual differences in incidental pitch memory. Jakubowski and Müllensiefen (2013) found that relative pitch performance (i.e., judging whether a short melody was the same or different as a version played immediately beforehand) was related to incidental pitch memory, measured through vocally producing familiar songs, suggesting a role of short-term auditory memory in long-term pitch memory for familiar songs. Subsequent work has demonstrated that a simple assessment of short-term tonal memory (i.e., adjusting a starting tone to match an initially presented target tone) explains variance in incidental pitch memory, even when controlling for factors such as song familiarity, fluid intelligence, musical training, singing accuracy, and auditory imagery (Van Hedger et al., 2018, 2024).

While the discussed findings suggest a consistent role of auditory short-term memory for both explicitly acquired pitch labels and incidentally acquired pitch memory, the relationship between these three factors has not been tested in the same participants, making it unclear how these constructs might influence one another. The relationship between incidental pitch memory and explicit pitch labeling is theoretically motivated (Levitin & Rogers, 2005) and is to some degree self-evident, as it is difficult to imagine good pitch labeling without a foundation of good pitch memory. Yet, surprisingly, the relationship between pitch memory and pitch labeling has primarily been addressed among AP possessors. Specifically, prior research has found that AP possessors have an enhanced incidental memory for absolute pitch (Dooley, 2011), even in situations where their explicit category labels would not be helpful (Van Hedger et al., 2016). These findings suggest that AP might develop in part due to enhanced pitch encoding and memory independent of category labeling (Ross et al., 2003, 2005); however, the extent to which individual variability in incidental pitch memory relates to the explicit acquisition of pitch labels among a non-AP sample, particularly when also considering auditory short-term memory, is unclear.

One possibility is that incidental pitch memory would explain significant variance in the explicit acquisition of AP labels, even when controlling for auditory short-term memory. This possibility is motivated by the fact that both incidental pitch memory and pitch labeling are grounded in long-term memory for pitch chroma and the relationship between pitch memory and pitch labelling is a core component of the two-stage model of AP (Levitin & Rogers, 2005). However, it is also possible that auditory short-term memory might be the driving force in explaining variance in the explicit acquisition of AP labels. This possibility is motivated by the fact that auditory short-term memory appears to relate to both incidental pitch memory and explicit pitch labeling even when controlling for a variety of factors, including musical training, tonal language experience, and singing accuracy, with auditory short-term memory significantly mediating the relationship between early musical training and explicit AP category labeling (Van Hedger et al., 2015), as well as mediating the relationship between singing accuracy and incidental pitch memory (Van Hedger et al., 2024). As such, auditory short-term memory appears to play a foundational role in the expression of both incidentally and explicitly acquired representations based on absolute pitch. However, addressing these possibilities requires a design in which incidental pitch memory, auditory short-term memory, and explicit pitch labeling are assessed within the same individuals.

Using a preregistered, within-participants design, the present study assessed how auditory short-term memory related to both incidental and explicit absolute pitch. Based on the two-component model of AP (Levitin & Rogers, 2005), we predicted that incidental pitch memory for popular songs would relate to how well individuals were able to learn explicit AP category labels in a rapid learning assessment. Furthermore, based on the reported associations between auditory short-term memory and both incidental pitch memory (Jakubowski & Müllensiefen, 2013; Van Hedger et al., 2018, 2024) and explicit AP category learning (Van Hedger et al., 2015, 2019; Wong, Ngan et al., 2020b), we predicted that auditory short-term memory would mediate the relationship between incidental pitch memory and explicit AP category learning, even when controlling for relevant factors to pitch labeling such as prior musical training and tonal language experience.

Method

Participants

A total of 192 individuals participated in the study and 179 participants were retained for analysis ($M = 27.42$ years old, $SD = 5.53$, range of 17 to 38 years old; 86 men, 88 women, 3 third gender/non-binary, 2 prefer not to answer). Ten participants were excluded from analysis because they failed more than one of six auditory attention checks, distributed throughout the study (see Procedure). An additional two participants were excluded because their scores on the AP Generalization Test were near ceiling (93.8% and 83.3%) and were demonstrably higher than the next highest score (50%), suggesting that they might have possessed AP, even though these participants did not disclose this. One participant was excluded because they recognized fewer than 50% of the popular songs used in the absolute pitch memory task. Participants were recruited from two primary sources: (1) Amazon Mechanical Turk ($n = 151$), and (2) an undergraduate participant pool ($n = 41$). The Mechanical Turk participants were further recruited through CloudResearch (Litman et al., 2017).

Participants were not specifically recruited based on their musical training, and the sample was largely musically untrained. Only 70 of 179 analyzable participants (39.1%) reported any prior musical training. The musically trained individuals reported a mean of 6.11 years (*SD*: 5.01 years) of musical training; however, the vast majority ($n =$ 53, 75.7%) of these participants reported that they were no longer actively playing music. Of the 17 participants who reported that they were still actively playing music, nine reported playing music for less than one hour per week, six reported playing music between one and three hours per week, and two reported playing music for more than five hours per week.

The research was approved by the Huron Research Ethics Board (Protocol #16S-202312) and all participants were treated in accordance with the Declaration of Helsinki. The study protocol was additionally preregistered (AsPredicted #156990). Although the preregistration included a sample size specification, it did not specifically address statistical power. However, as calculated via G*Power (Faul et al., 2009), the achieved sample size ($n = 179$) was adequately powered ($1 - \beta = .80$) to detect small-to-medium sized correlations ($r = .21$), which we considered adequate given the previously reported association strengths between tonal short-term memory and incidental absolute pitch (e.g., $r =$ -.43 and $r = .31$ in Van Hedger et al., 2018 and $r = -.38$ in Van Hedger et al., 2024), as well as between tonal short-term memory and explicit AP categorization (e.g., $r = -.66$ in Van Hedger et al., 2015).

Materials

The Tonal Short-Term Memory (STM) task was modeled on prior work (Heald et al., 2014; Van Hedger et al., 2018, 2024) and involved adjusting a starting note to match a target note. On each trial participants first heard a target note (either F#, G, G#, or A) by clicking on a designated button on the computer screen. This target note was followed by 1000 ms of masking noise, and participants could not replay the target note for the remainder of the trial. Participants then clicked on a second button to hear a starting note. Starting notes could be any of the remaining pitch chroma categories (i.e., A#, B, C, C#, D, D#, E, or F). After listening to the starting note, participants attempted to match the target note by clicking on "up" and "down" arrows displayed on the screen, which increased or decreased the starting tone by one-third of a semitone, respectively. All tones were 250 ms in duration. Unlike previous implementations of the task, we used Shepard tones to better represent short-term memory for pitch chroma rather than pitch height. This is because Shepard tones are built by stacking sine waves separated by octaves in a manner that renders them ambiguous in terms of pitch height (Shepard, 1964), despite possessing a clear pitch chroma (e.g., Deutsch, 1986). The tonal short-term memory task had two practice trials and 32 scored trials (4 target notes fully crossed with 8 starting notes in a randomized order). Performance was operationalized as the mean absolute deviation from the target note, represented in terms of arrow clicks (e.g., a mean score of 7 would reflect that a participant was, on average, 2.33 semitones away from the target note). All responses in the Tonal STM were untimed (i.e., participants had an unlimited amount of time to respond).

The Incidental AP Task required participants to judge whether short (5000 ms) excerpts from musical recordings had been shifted in pitch. The 20 excerpts (Appendix 1) were a subset of the 28 excerpts used in recent work (Van Hedger et al., 2024), which had been selected based on high levels of familiarity. There were 10 correct (i.e., unshifted) and 10 incorrect (i.e., shifted) trials, presented in a random order. Five of the incorrect excerpts were shifted by +1 semitone and the other five were shifted by −1 semitone. Correct stimuli were also shifted in pitch (first up by 0.5 semitones and then down by 0.5 semitones) to minimize the use of audio quality cues introduced by the pitch shifting procedure. Pitch shifting was done in Audacity using a "high-quality stretching" option to preserve the length of each recording. The assignment of an excerpt to be either correct, incorrect by +1 semitone, or incorrect by −1 semitone was randomized across participants. Participants made a forced-choice judgment whether the excerpt sounded correct. Participants were also asked to rate their familiarity with the song and their confidence in their pitch judgment on a Likert-type scale, ranging from 0 (*Not at all*) to 4 (*Extremely*). Trials were discarded if participants reported no prior familiarity with the recording, and participants were discarded from subsequent analysis if they did not recognize at least 50% of the recordings. In the present study, only one participant was removed for recognizing fewer than 50% of the recordings, and of the remaining participants, only 2.6% of trials were discarded due to no reported familiarity. Incidental AP performance was operationalized in terms of proportion correct for the forced-choice judgment about whether the recording sounded correct. All responses in the Incidental AP Task were untimed (i.e., participants had an unlimited amount of time to respond).

In the AP Categorization Task, participants learned to associate musical notes with arbitrary labels. The task was modeled on prior AP learning work, in which participants learned to associate pitch chroma categories with names (Gervain et al., 2013). Participants were trained on a total of six notes (C, D, E, F#, G#, A#), which were associated with six names (Sarah, David, Francine, Jimmy, Karen, Leo). The musical notes used in training were 1000 ms in duration and consisted of multiple instruments simultaneously playing the note across multiple octaves. This was done to encourage participants to attend to pitch chroma, as opposed to pitch height in making their categorization judgments (cf. Bongiovanni et al., 2023). On each trial, participants would hear a note and would then be presented with six buttons. Participants then made a forced-choice judgment as to who played the note. During training (108 trials: 18 repetitions of each note in a randomized order), participants would receive visual and auditory feedback to facilitate learning. The visual feedback circled the correct button and printed on the screen "[Correct! / Incorrect.] That note was played by [Name]." During this visual feedback, participants reheard the note. Participants did not receive any initial instructions or pre-training trials that specified which sound was associated with each label, meaning that the initial trials of training inherently involved guessing and attending to the provided feedback. Following training, participants completed two tests of AP learning – a Specific Test, which was identical to training with the exception that participants did not receive feedback, and a Generalization Test, which also did not provide feedback to participants and additionally used Shepard tones to assess timbre generalization and further discourage the use of pitch height as a cue. Both tests consisted of 48 trials (8 repetitions of each note in a randomized order). All responses in the AP Categorization Task were untimed (i.e., participants had an unlimited amount of time to respond).

The questionnaire consisted of basic demographic, musical experience, and language experience questions. The demographic questions consisted of age, gender, and highest level of education obtained. The musical experience questions included whether participants thought they possessed perfect pitch (*Yes*, *No*, *Not sure*), whether participants had any musical training (*Yes, No*), and if *Yes*, (1) at what age they began musical instruction, (2) what they considered their primary instrument, (3) how many years they play(ed) music, and (4) how many hours per week they currently play music. The language experience questions included whether participants considered English to be their native language, and if not, to provide their native language in a provided textbox.

The study was programmed in jsPsych 7 (De Leeuw, 2015; De Leeuw et al., 2023). All sounds were RMS normalized to −15 dB relative to full scale and were presented at 44.1 kHz with 16-bit depth. Participants who completed the study in person wore either Sony MDR-7506 or Audio-Technica AT-M20x studio monitor headphones and completed the study in a quiet, designated testing space.

Procedure

All participants, regardless of completing the study online or in person, completed the study using the same study link. The study link used a URL variable to differentiate participants who completed the study online versus those who completed the study in person, and this variable was logged to the data output for analysis.

After reading the letter of information and providing informed consent via the computer, participants completed a short auditory calibration. The auditory calibration involved a volume adjustment, during which participants heard a sample of music (normalized to the same level as the other auditory stimuli used in the study) and adjusted the computer's volume so that the sample of music was heard at a comfortable volume. Following the auditory calibration, participants were asked whether they were wearing headphones or earbuds (*Yes*, *No*). Following the self-report headphone question, participants completed a short perceptual task based on Milne et al. (2021) that was designed as a performance-based measure of whether participants were wearing headphones or earbuds. The task (6 trials) involves listening to three samples of noise on each trial and indicating which of the three noises contains a tone. The tone is only perceptible when presented dichotically, and a correct answer on at least 5 of 6 trials was taken to indicate headphone use. Although the auditory calibration could be viewed as largely unnecessary for the in-person participants, we decided to include it to (1) keep the procedure identical across samples, and (2) to assess the relationship between self-reported and performance-based headphone use among a sample that was known to be wearing headphones. These results are not formally reported here; however, the data are available through Open Science Framework. Given that headphone use was encouraged but not required for the online participants, the results of the headphone assessment were not used to exclude any participants. Moreover, exploratory analyses suggested that headphone use was not a significant predictor of AP Categorization, nor did its inclusion change the interpretation of any other predictor variables (also reported on Open Science Framework). The auditory calibration took less than five minutes to complete.

Following the auditory calibration, participants completed the Tonal STM Task, the Incidental AP Task, the AP Categorization Task, and the questionnaire in this order. The Tonal STM Task took approximately 10 minutes to complete, the Incidental AP Task took approximately 10 minutes to complete, the AP Categorization Task took approximately 25 minutes to complete, and the questionnaire took less than five minutes to complete. We opted to fix the ordering of the pitch memory tasks because we wanted to present the AP Categorization Task last, as this was the only pitch memory task that provided feedback and thus could have ostensibly influenced performance on the other pitch memory assessments depending on the extent of learning. Although this still could have allowed us to vary whether the Incidental AP Task or the Tonal STM task was presented first, prior work has found comparable associations between these tasks regardless of order (e.g., Experiment 2 of Van Hedger et al., 2018, 2024). In total, the study took approximately 50 minutes to complete.

Each of these three main auditory tasks (Tonal STM Task, Incidental AP Task, AP Categorization Task) contained two auditory attention checks – one embedded within the first half of the trials, and one embedded within the second half of the trials. In the case of the AP Categorization Task, the attention checks were embedded within the training phase of the task. Each attention check prompted participants via an auditory cue to press the 'spacebar' within 10 seconds. No visual cues were provided during these attention checks. If a participant failed to respond within 10 seconds, the study moved on automatically and the trial was marked as a "timeout." Participants were discarded from the main analysis if they logged more than one timeout.

Data analysis

All analyses were performed in R version 4.3.1 using RStudio version 2023.12.1.402. To assess whether participants were above chance in the Incidental AP Task and in the AP Categorization Task, we used one-sample *t*-tests against chance performance. Chance performance was 50% in the case of the Incidental AP Task (as it was a two-alternative forced choice response), whereas chance performance was 16.7% in the case of the AP Categorization Task (as it was a six-alternative forced choice response). Performance differences across the three blocks of the AP Categorization Task were additionally compared using paired-samples *t*-tests (e.g., to assess the extent to which performance was attenuated in the testing blocks). All *t*-tests report Cohen's d effect sizes, calculated via the "effsize" package v.0.8.1 (Torchiano, 2016).

Relationships among the Tonal STM Task, the Incidental AP Task, and the AP Categorization Task were first assessed using Pearson correlations. We then constructed multiple regression models, using performance on the AP Categorization Task as the dependent variable, to determine how performance on the Tonal STM Task and Incidental AP Task related to AP categorization performance. Separate models were created for each phase of the AP Categorization Task (Training, Specific Test, Generalization Test). Each model contained Tonal STM performance, Incidental AP performance, reported musical training (yes, no), whether participants actively play music (yes, no), tonal language proficiency (yes, no), and whether the participant was run online (yes, no) as predictor variables. Although prior AP training research has used the age of beginning musical instruction as a measure of musical training (Van Hedger et al., 2015), in the present study we opted to use the categorical distinction of whether or not a participant reported any musical training. This was largely because only 39.1% of participants reported musical training in the present study, and thus calculating the age of beginning musical instruction would be problematic for most participants in the present sample. Given that AP Categorization performance was a proportion, we used generalized linear models with a quasibinomial link. Each model was weighted based on the total number of trials (108 for the Training model, and 48 for both the Specific and Generalization Test models).

To test whether Tonal STM mediated the relationship between Incidental AP performance and AP Categorization performance, we used the "mediation" package (Tingley et al., 2014). The average causal mediation effect and direct effect were estimated through a bootstrapping procedure (5,000 simulations). The significance of each effect was determined by assessing whether the 95% confidence interval from the bootstrapping procedure included zero.

Deviations from preregistration

There were some minor deviations from preregistration related to participant recruitment. The preregistration document specified a sample size of 200, with 160 participants coming from Amazon Mechanical Turk and 40 participants coming from in-person testing using a university student sample. The Amazon Mechanical Turk sample was smaller than expected because a small minority of respondents did not complete the study and provided an incorrect (duplicate) completion code, leading to their work being rejected. Additionally, although we reached our planned sample size for the university sample, a small subset of these participants (n = 7) completed the study online, similar to the Amazon Mechanical Turk participants. As such, all models include a term indicating whether participants completed the study in a supervised versus unsupervised setting.

We also deviated from the preregistered exclusion criteria with respect to AP performance. In preregistration, we state that a participant would be excluded if they self-reported possessing AP *and* scored above two standard deviations from the mean in either the absolute pitch memory task or the AP categorization task. The two excluded participants who scored close to ceiling (93.8% and 83.3%) on the AP Generalization Test were indeed over two standard deviations higher than the mean (6.01 and 5.15 standard deviations above the mean, respectively); however, they did not identify as AP possessors. Nevertheless, given how much these two data points differed from the rest of the sample, we opted for a more conservative approach in excluding these participants, to minimize the potential for influential points in our regression analyses.

Although there were no deviations from the planned analyses, the preregistration plan did not include direct comparisons of performance across the three blocks of the AP Categorization Task (e.g., testing the degree to which performance was attenuated in the Generalization Test relative to the Specific Test). Thus, these analyses should be considered exploratory. Additionally, the Pearson correlations among the regression variables, reported in Table 1, were added during peer-review to provide greater clarity and transparency in interpreting the regression output. These correlations were therefore not preregistered and should also be considered exploratory. There were no deviations to preregistration with respect to the materials or the procedure.

Results

Comparing absolute pitch assessments to chance performance

Participants in the Incidental AP Task correctly categorized the popular song excerpts 59.8% of the time (SD = 12.1%), which was significantly above the chance estimate of 50% with a large effect size, $t(178) = 10.81$, $p < .001$, $d = 0.81$. In the AP Categorization Task, participants accurately labeled 50.1% of the notes during Training (SD = 17.0%), which was significantly above the chance estimate of 16.7% with a very large effect size, $t(178) = 26.21$, $p < .001$, $d = 1.96$. Although this finding is not surprising given that participants received feedback during training, it suggests that participants were engaged with the AP Categorization Task, and it additionally provides a baseline against which the Specific and Generalization Tests – which provided no feedback to participants – can be compared.

Above-chance performance was also observed in the Specific Test, despite participants no longer receiving feedback. Mean performance in the Specific Test was 41.9% (SD = 17.1%), which was significantly above the chance estimate of 16.7% with a very large effect size, $t(178) = 19.68$, $p < .001$, $d = 1.47$. However, it should be noted that performance in the Specific Test was significantly attenuated compared to

Table 1 Pearson correlations among predictor and outcome variables used in the multiple regression models

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. IAP	-								
2. TSTM	-.19*	-							
3. Musician	.19*	-.11	-						
4. Active	.06	-.06	.40***	-					
5. Tonal	-.03	.01	-.04	-.05	-				
6. Training	.23**	-.66***	.13	-.10	.02	-			
7. Specific	.24**	-.52***	.17*	-.03	-.05	.75***	-		
8. Generalization	.09	-.34***	.15*	.10	.09	.34***	.41***	-	
9. Online	-.21**	-.05	.00	.04	-.12	-.02	.02	.05	

IAP = Incidental AP; TSTM = Tonal Short-Term Memory; Musician = Reported Musical Training (binary); Active = Actively Playing Music (binary); Tonal = Tonal Language Speaker (binary); Training = AP Categorization Training; Specific = Specific Test AP Categorization; Generalization = Generalization Test AP Categorization; Online = Whether participants completed the study in a monitored or unmonitored environment (binary). Values represent uncorrected Pearson correlations. *** $p < .001$ ** $p < .01$ * $p < .05$

Training, $t(178) = -9.03$, $p < .001$, $d = 0.48$. Performance in the Generalization Test was attenuated relative to performance in the Specific Test, $t(178) = -18.09$, $p < .001$, $d = 1.47$, suggesting that participants were only partially able to generalize learning to a new timbre that discouraged the use of pitch height. Nevertheless, mean performance in the Generalization Test was 20.4% ($SD = 9.8\%$), which was still significantly above the chance estimate of 16.7% with a small-to-medium effect size, $t(178) = 5.10$, $p < .001$, $d = 0.38$. Figure 1 plots performance on the AP Categorization Task.

Correlations across pitch measures

There were several significant correlations among our pitch measures, as predicted. First, Tonal STM performance was significantly correlated with Incidental AP performance, $r(177) = -.19$, $p = .011$, such that better Tonal STM performance (i.e., a lower deviation score) was associated with better Incidental AP performance (i.e., a larger proportion of correct trials). This association conceptually replicates prior work (Van Hedger et al., 2018, 2024). Second, the blocks of the AP Categorization Task were all significantly and positively intercorrelated. Performance on AP Categorization Training was positively correlated with both performance on the Specific Test, $r(177) = .75$, $p < .001$, as well as performance on the Generalization Test, $r(177) = .34$, $p < .001$, and performance on the Specific Test was positively correlated with performance on the Generalization Test, $r(177) = .41$, $p < .001$. Third, Tonal STM was significantly associated with performance on the AP Categorization Task (Fig. 2, top), such that better Tonal STM performance was associated with better AP categorization across all blocks (Training, Specific Test, Generalization Test). Finally, performance on the Incidental AP Task was associated with performance on the AP Categorization Task, at least for the Training and Specific Test blocks (Fig. 2, bottom). In contrast, the association between performance on the Incidental AP Task and the Generalization Test was nominally in the predicted direction but was not significant. In sum, the three assessments of pitch memory showed consistent intercorrelations in the predicted direction (i.e., better performance in one pitch task associated with better performance in another pitch task).

Modelling absolute pitch categorization

Simple correlations among variables used in the regression models are printed in Table 1, and the model results are printed in Table 2. Both Tonal STM and current active musical playing were significant predictors in the model examining AP categorization performance in Training. The relationship between Tonal STM and AP categorization performance was in the predicted direction (i.e., better Tonal STM associated with better AP categorization performance); however, the relationship between active musical playing and AP categorization performance in Training was negative, meaning individuals who were actively playing music performed *worse* compared to individuals who were not actively playing music. However, the simple association between actively playing music and AP categorization performance in Training was not significant (Table 1), suggesting that active musical training was operating as a suppressor variable in the model. Both Incidental AP performance and Tonal STM performance were significant predictors in the model examining AP categorization performance in the Specific Test in the predicted direction (i.e., better Tonal STM and better Incidental AP associated with better AP categorization performance). Finally, Tonal STM was the only significant predictor in the model examining AP categorization performance in the Generalization Test in the

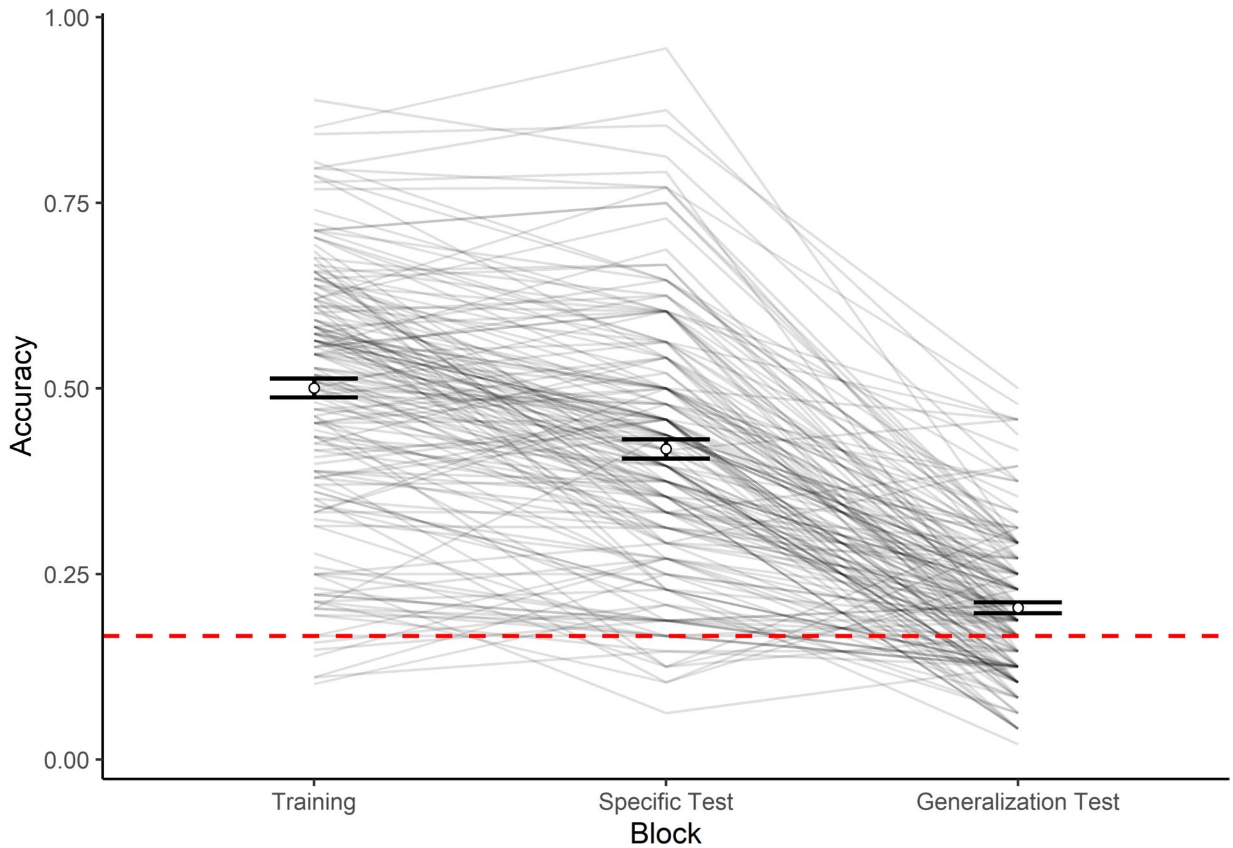

Fig. 1 Training and testing performance for the AP categorization task. *Note*: Error bars represent plus or minus one standard error of the mean. Filled circles represent mean performance for each block. Grey lines represent individual participant values. The dashed red line represents chance performance in the task

predicted direction (i.e., better Tonal STM associated with better AP categorization performance). Thus, Tonal STM was the only consistent predictor of performance in the AP Categorization Task across all blocks, even when controlling for sample (online versus in-lab), tonal language proficiency, and musical factors.

Mediation analyses

There was evidence that Tonal STM fully mediated the relationship between Incidental AP and AP categorization performance in Training. The average causal mediated effect was 0.06, with the 95% confidence interval not including zero [0.01, 0.11]. The proportion mediated was 0.60, with the 95% confidence interval also not including zero [0.17, 1.20]. In contrast, the direct effect of Incidental AP on Training performance was 0.04, with the 95% confidence interval including zero [−0.01, 0.10]. The total effect (i.e., the combination of both the direct and indirect effects – here, the effect of Incidental AP on Training performance without considering Tonal STM) was 0.10, with the 95% confidence interval not including zero [0.03, 0.18].

For the Specific Test, there was evidence that Tonal STM partially mediated the relationship between Incidental AP and AP categorization performance. The average causal mediated effect was 0.08, with the 95% confidence interval not including zero [0.02, 0.14]. The proportion mediated was 0.41, with the 95% confidence interval also not including zero [0.11, 0.97]. The direct effect of Incidental AP on Specific Test performance was 0.11, with the 95% confidence interval not including zero [0.002, 0.24], suggesting that Incidental AP still contributed to explaining variance in Specific Test performance. The total effect of Incidental AP on Specific Test performance (without considering Tonal STM) was 0.18, with the 95% confidence interval not including zero [0.06, 0.33].

The results from the Generalization Test should be interpreted with caution, as the association between Incidental AP and AP categorization performance was not significant. Nevertheless, we decided to assess evidence for mediation given our preregistration. In this analysis, the average causal mediated effect was 0.05, with the 95% confidence interval not including zero [0.005, 0.08]. The proportion mediated was 0.80, although the 95% confidence

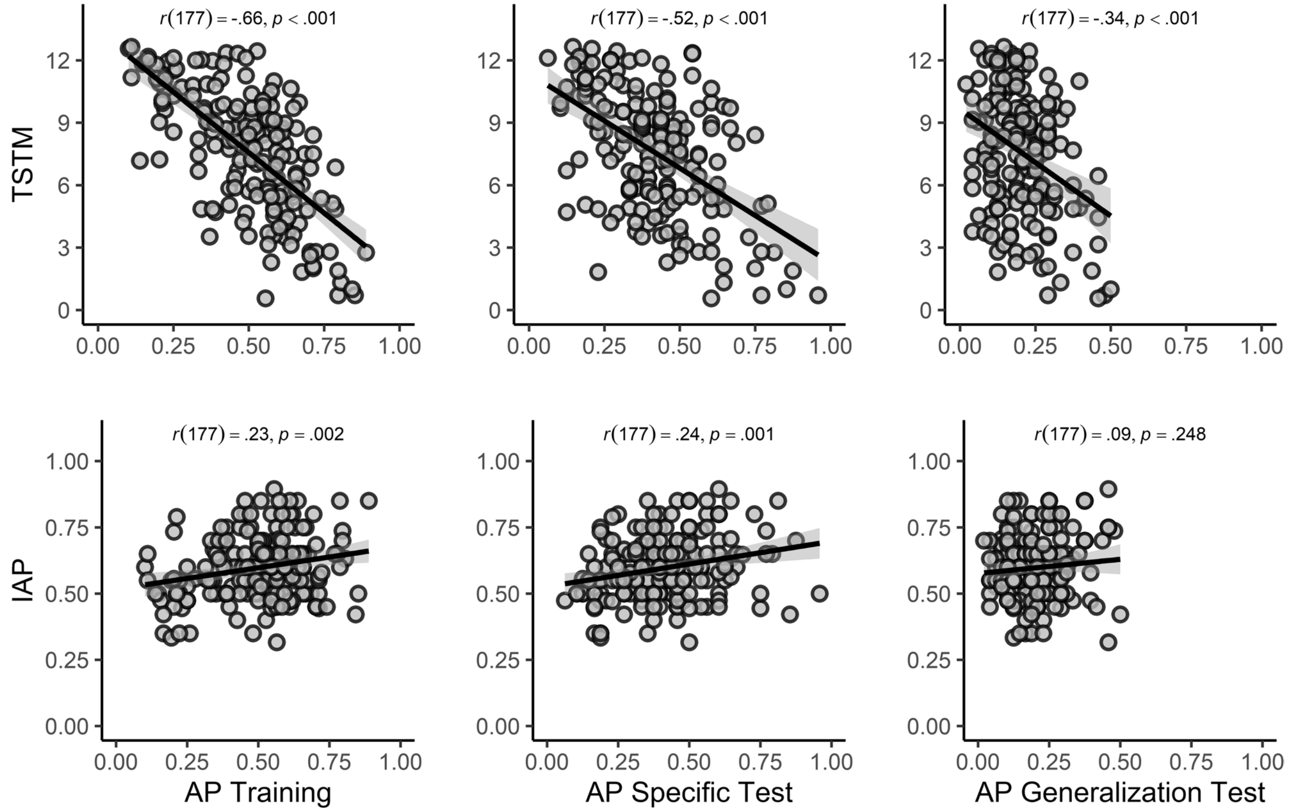

Fig. 2 Correlations between Tonal Short-Term Memory (TSTM) and AP Categorization (**top row**), and Incidental AP (IAP) and AP Categorization (**bottom row**). *Note:* TSTM = Tonal Short-Term Memory Task; IAP = Incidental AP Task. Error ribbons represent 95% confidence intervals. Each column represents one of the three blocks of the AP Categorization Task (Training, Specific Test, and Generalization Test)

Table 2 Summary of multiple regression models explaining variance in the AP categorization task

Predictor	Training	Specific Test	Generalization Test
Online (Yes)	−0.05 (0.07)	0.02 (0.08)	0.06 (0.07)
Tonal STM	−0.10 (0.01) ***	−0.07 (0.01) ***	−0.04 (0.01) ***
Incidental AP	0.32 (0.22)	0.50 (0.25) *	0.04 (0.21)
Music Training (Yes)	0.11 (0.06)	0.12 (0.06)	0.07 (0.05)
Actively Playing Music (Yes)	−0.30 (0.09) **	−0.19 (0.11)	0.05 (0.09)
Tonal Language (Yes)	0.06 (0.17)	−0.12 (0.20)	0.23 (0.16)
Pseudo R^2	.47	.30	.14

Values represent unstandardized beta coefficients. Standard errors are printed in parentheses. *** $p < .001$ ** $p < .01$ * $p < .05$

interval contained zero [−5.35, 6.34]. The direct effect of Incidental AP on Generalization Test performance was 0.01, with the 95% confidence interval containing zero [−0.10, 0.13]. Finally, the total effect of Incidental AP on Generalization Test performance (without considering Tonal STM) was 0.06, with the 95% confidence interval containing zero [−0.07, 0.17]. The results from the mediation analyses are plotted in Fig. 3.

Discussion

The present study examined the relationship among different assessments of absolute pitch memory, differing both in terms of timescale (short- versus long-term memory) and memory encoding (incidental versus intentional). Despite these differences among the pitch memory

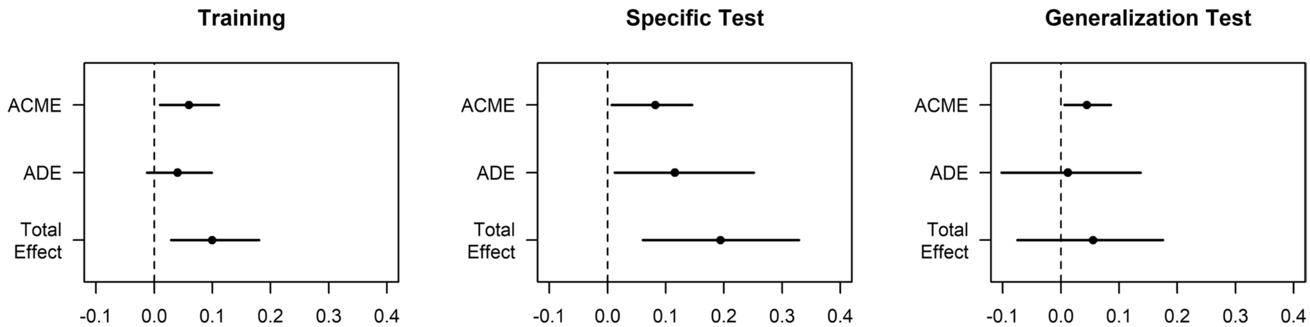

Fig. 3 Mediation analyses for each block of the AP categorization task. *Note*: Each mediation model also included sample (in-lab versus online), musical training (yes, no), whether participants actively played music (yes, no), and tonal language proficiency (yes, no) as control variables. ACME = Average Causal Mediated Effect, ADE = Average Direct Effect. Lines represent 95% confidence intervals generated from bootstrapping (5,000 samples)

assessments, we predicted that both incidentally acquired AP (e.g., remembering the absolute pitches of familiar songs without intention to do so) and Tonal STM (e.g., precisely remembering a pitch on the order of seconds) would both relate to one's ability to explicitly learn categories based on pitch chroma, even when controlling for other relevant musical and linguistic factors.

Our results largely supported these predictions. Both Incidental AP and Tonal STM were positively correlated with AP categorization performance during initial Training (with feedback) and in the Specific Test (same sounds as training with no feedback). In the Generalization Test, only Tonal STM was significantly correlated with AP categorization performance. Regression models and mediation analyses further supported the foundational role of Tonal STM in explaining variance in AP categorization performance. Tonal STM was the only consistent significant predictor of AP categorization performance across Training, the Specific Test, and the Generalization Test, even when controlling for participant sample, musical factors, tonal language experience, and Incidental AP. Furthermore, Tonal STM fully mediated the relationship between Incidental AP and AP categorization performance in Training, and Tonal STM partially mediated the relationship between Incidental AP and AP categorization performance in the Specific Test. These results align with a growing body of research that has associated auditory short-term and working memory to pitch chroma representations (e.g., Deutsch & Dooley, 2013; Van Hedger et al., 2015, 2018, 2019, 2024; Van Hedger & Nusbaum, 2018; Wong, Ngan et al., 2020b).

The present study helps inform the theoretical understanding of AP in two primary ways. First, the present results can be interpreted within the influential two-step model of AP (e.g., see Levitin & Rogers, 2005). Although Levitin and Rogers (2005) suggest that good pitch memory is likely a prerequisite for pitch labelling, this association has received relatively little attention from prior research. One exception comes from Jakubowski and Müllensiefen (2013), who assessed the relationship between incidental pitch memory and a pitch decay task, which was used as a proxy for pitch labelling ability. In this work, the authors did not find a significant association between incidental pitch memory and the pitch decay task. However, one potential explanation for this null result is that the pitch decay task was meant to tap into existing labelling ability (i.e., it was not focused on the trainability of labels), with the authors reporting that non-AP possessors in a pilot study performed at chance on the decay task. The current findings, by focusing on the rapid trainability of labels based on pitch chroma, are consistent with the conjecture by Levitin and Rogers (2005) that pitch memory is associated with pitch labelling; however, the current results additionally suggest that this relationship is driven by short-term memory for pitch. Thus, the current results support the two-step model of AP but emphasize the importance of short-term pitch memory in both the incidental formation of absolute pitch memory (e.g., from popular songs) as well as the explicit learning and memory of categories based on pitch chroma.

The second way in which the current results inform the theoretic understanding of AP is by suggesting that AP be conceptualized along a continuum, rather than as a dichotomous ("all-or-none") ability (cf. Athos et al., 2007). In the present study, we observed significant variability in AP category learning, with some individuals scoring over 75% in the Specific Test and over 50% in the Generalization Test, and other participants scoring at or even below chance (16.7%) on these same assessments (see Fig. 1). Although the learning we observed in the present study is still below what would typically be considered a threshold for the strictest thresholds of AP (e.g., Baharloo et al., 1998), an increasing body of research that has examined variability in AP performance has shown that where one draws the line between AP and non-AP is often arbitrary (Bermudez & Zatorre, 2009; Van Hedger et al., 2020). As such, the finding that Tonal STM in the present study is just as strongly associated with the generalization of learned AP categories (r = -.34) as how well one performs in the training block of AP categorization *in the same task* (r = .34) emphasizes the role of short-term memory for pitch in the explicit learning of categories based on pitch chroma. This finding also aligns with the discussed role of (enhanced) auditory short-term memory among AP possessors (Deutsch & Dooley, 2013; Van Hedger & Nusbaum, 2018), and further supports the notion that the ability to label pitches based on chroma is a continuously distributed ability that critically relates to auditory short-term memory.

Although the present results largely fit within a broader theoretical framework of AP, there were some unexpected results, such as the nonsignificant correlation between performance on the Incidental AP Task and performance on the Generalization Test of the AP Categorization Task. However, this null association might be best explained when considering the multidimensional nature of pitch – specifically, considering the difference between pitch *height* and pitch *chroma*. The Generalization Test used Shepard Tones, which are designed to have a clear pitch chroma but be ambiguous with respect to pitch height (Shepard, 1964). In contrast, the Incidental AP Task involved shifting excerpts of popular recordings up or down in pitch, which alters both pitch height and pitch chroma. Thus, while the Incidental AP Task is clearly tapping into a long-term memory for pitch, it is less clear from the current task whether this long-term memory is driven by long-term memory for pitch height. More recent work has shown that incidental absolute pitch memory is still observable in contexts where familiar melodies are shifted in octave and played in a different timbre (Van Hedger et al., 2023). Thus, future work might consider redesigning the Incidental AP Task to more strongly encourage the use of pitch chroma, to assess whether this would strengthen the association with the Generalization Test.

The present study has some limitations that influence the generalizability of the findings. First, the present study trains AP categories in a single session of learning. As such, the improvements seen in the present study with respect to AP categorization were modest (particularly within the Generalization Test) and are not meant to comment on the trainability of "genuine" AP ability. Indeed, given that "genuine" AP possessors commonly report that the process of acquiring AP is not deliberate or effortful (Deutsch, 2013), this might suggest that the cognitive mechanisms discussed here might not underly the rapid acquisition of "genuine" AP in childhood. However, given conceptualizations of AP as existing along a continuum (e.g., Bermudez & Zatorre, 2009), and given that previous single session approaches to training AP have found stable performance six months after training (Van Hedger et al., 2015), we still consider the present research findings to be relevant to the understanding of absolute pitch memory in humans more generally and to the phenomenon of AP more specifically. Second, the use of single tasks to measure constructs (e.g., a single task to measure Tonal STM, a single task to measure Incidental AP) makes it difficult to know whether the *constructs* of Tonal STM and Incidental AP, as opposed to the tasks used to measure these constructs, are ultimately important in explaining variance in AP category learning. The use of single tasks to measure constructs was done for the sake of time efficiency given the within-participant design; however, future work should consider a more comprehensive approach (e.g., creating assessment batteries for a particular construct and using a latent variable approach) to replicate the current findings. Third, it is possible that the participants who completed the study in an unmonitored setting used external aides (e.g., piano keyboards) to augment their pitch memory results. However, we find this explanation unlikely for several reasons, most notably because the online and in-lab participants did not differ in their pitch memory accuracy scores. Nevertheless, future studies could address this by including a limited window of time within which participants can respond.

In conclusion, the present study assessed how short-term memory for pitch and incidental absolute pitch memory related to AP category learning. The findings suggest a foundational role of short-term memory for pitch in explaining variance in AP category learning, which aligns with prior work in the AP learning literature (Van Hedger et al., 2015), the incidental absolute pitch memory literature (Van Hedger et al., 2018), and the "genuine AP" literature (Deutsch & Dooley, 2013). Although incidental absolute pitch memory was associated with some aspects of explicit AP category learning, which has been proposed (Levitin & Rogers, 2005) but not directly tested to our knowledge, this association was mediated by short-term memory in the present study. Taken together, these findings highlight short-term memory for pitch as a critical individual difference measure in understanding the formation and maintenance of perceptual categories based on pitch chroma, which has implications for AP and beyond.

Appendix

Title	Artist
Single Ladies	Beyonce
Umbrella	Rihanna
Shake It off	Taylor Swift
Toxic	Britney Spears
Rolling in the Deep	Adele
Firework	Katy Perry
Blinding Lights	The Weeknd
Hips Don't Lie	Shakira
Call Me Maybe	Carly Rae Jepsen
Uptown Funk	Bruno Mars
Poker Face	Lady Gaga
Starships	Nicki Minaj
Royals	Lorde
Party in the USA	Miley Cyrus
Happy	Pharrell Williams
Despacito	Luis Fonsi
Gangnam Style	Psy
Take On Me	a-ha
We Are The Champions	Queen
Somebody That I Used to Know	Gotye

List of the 20 popular song excerpts used in the Incidental AP Task

Authors' contributions **SVH**: conceptualization, methodology, software, formal analysis, data curation, writing – review & editing, supervision, funding acquisition. **KJ**: methodology, investigation, writing – review & editing, project administration. **AG**: conceptualization, methodology, writing – review & editing, **SYH**: methodology, investigation, writing – review & editing, project administration.

Funding This study was supported through a SSHRC Institutional Grant awarded to SVH by the Centre for Undergraduate Research Learning at Huron University.

Availability of data and materials All data and study materials are available through Open Science Framework (https://doi.org/10.17605/OSF.IO/EDAJS)

Code availability The study code (both the code to run the study as well as the code to analyze the data) is available through Open Science Framework (https://doi.org/10.17605/OSF.IO/EDAJS)

Declarations

Conflicts of interest The authors declare no conflicts of interests.

Ethics approval This research was approved by the Huron University Research Ethics Boards (#16S-202312).

Consent to participate All participants were provided with a Letter of Information, providing the details of the study, at the beginning and were provided a Debriefing Letter, explaining the purpose of the study, at the end. All participants provided consent to participate in the study.

Consent for publication Not applicable.

References

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, 104*(37), 14795–14800. https://doi.org/10.1073/pnas.0703868104

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics, 62*(2), 224–231. https://doi.org/10.1086/301704

Bermudez, P., & Zatorre, R. J. (2009). A distribution of absolute pitch ability as revealed by computerized testing. *Music Perception, 27*(2), 89–101. https://doi.org/10.1525/mp.2009.27.2.89

Bongiovanni, N. R., Heald, S. L. M., Nusbaum, H. C., & Van Hedger, S. C. (2023). Generalizing across tonal context, timbre, and octave in rapid absolute pitch training. *Attention, Perception, & Psychophysics, 85*(2), 525–542. https://doi.org/10.3758/s13414-023-02653-0

Crozier, J. B. (1997). Absolute pitch: Practice makes perfect, the earlier the better. *Psychology of Music, 25*(2), 110–119. https://doi.org/10.1177/0305735697252002

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

De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborativeecosystem of behavioral experiments. *Journal of Open Source Software, 8*(85), 5351. https://doi.org/10.21105/joss.05351

Deutsch, D. (1986). A musical paradox. *Music Perception, 3*(3), 275–280. https://doi.org/10.2307/40285337

Deutsch, D. (2013). Absolute Pitch. *The Psychology of Music* (pp. 141–182). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00005-5

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

Deutsch, D., Henthorn, T., & Dolson, M. (2004). absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception, 21*(3), 339–356. https://doi.org/10.1525/mp.2004.21.3.339

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America, 119*(2), 719. https://doi.org/10.1121/1.2151799

Dooley, K. (2011). *Absolute pitch and related abilities*. UC San Diego. ProQuest ID: Dooley_ucsd_0033D_11905. Merritt ID: ark:/20775/bb1537101t. Retrieved from https://escholarship.org/uc/item/7j46q4z5

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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

Gervain, J., Vines, B. W., Chen, L. M., Seo, R. J., Hensch, T. K., Werker, J. F., & Young, A. H. (2013). Valproate reopens critical-period learning of absolute pitch. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00102

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (1999). Absolute Pitch: Prevalence, Ethnic Variation, and Estimation of the Genetic Component. *The American Journal of Human Genetics, 65*(3), 911–913. https://doi.org/10.1086/302541

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

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

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

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences, 9*(1), 26–33. https://doi.org/10.1016/j.tics.2004.11.007

Levitin, D. J., & Zatorre, R. J. (2003). On the nature of early music training and absolute pitch: A reply to brown, sachs, cammuso, and folstein. *Music Perception, 21*(1), 105–110. https://doi.org/10.1525/mp.2003.21.1.105

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Lockhead, G. R., & Byrd, R. (1981). Practically perfect pitch. *The Journal of the Acoustical Society of America, 70*(2), 387–389. https://doi.org/10.1121/1.386773

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods, 53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Miyazaki, K., & Ogawa, Y. (2006). Learning absolute pitch by children. *Music Perception, 24*(1), 63–78. https://doi.org/10.1525/mp.2006.24.1.63

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. https://doi.org/10.1196/annals.1284.065

Ross, D. A., Gore, J. C., & Marks, L. E. (2005). Absolute pitch: Music and beyond. *Epilepsy & Behavior: E&B, 7*(4), 578–601. https://doi.org/10.1016/j.yebeh.2005.05.019

Russo, F. A., Windell, D. L., & Cuddy, L. L. (2003). Learning the "Special Note": Evidence for a critical period for absolute pitch acquisition. *Music Perception, 21*(1), 119–127. https://doi.org/10.1525/mp.2003.21.1.119

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

Schellenberg, E. G., & Trehub, S. E. (2008). Is There an Asian advantage for pitch memory? *Music Perception, 25*(3), 241–252. https://doi.org/10.1525/mp.2008.25.3.241

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America, 36*(12), 2346–2353. https://doi.org/10.1121/1.1919362

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin, 113*(2), 345–361. https://doi.org/10.1037/0033-2909.113.2.345

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

Torchiano, M. (2016). *Effsize—A package for efficient effect size computation* [Computer software]. Zenodo.https://doi.org/10.5281/ZENODO.1480624

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

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

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

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

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2019). Absolute pitch can be learned by some adults. *PLOS ONE, 14*(9), e0223047. https://doi.org/10.1371/journal.pone.0223047

Van Hedger, S. C., Veillette, J., Heald, S. L. M., & Nusbaum, H. C. (2020). Revisiting discrete versus continuous models of human behavior: The case of absolute pitch. *PLOS ONE, 15*(12), e0244308. https://doi.org/10.1371/journal.pone.0244308

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

Van Hedger, S. C., Halpern, A. R., Vollweiler, D. J., Smith, E. E., & Pfordresher, P. Q. (2024). Is Hey Jude in the right key? Cognitive components of absolute pitch memory. *Memory & Cognition*. https://doi.org/10.3758/s13421-024-01530-x

Wilson, S. J., Lusher, D., Martin, C. L., Rayner, G., & McLachlan, N. (2012). Intersecting factors lead to absolute pitch acquisition that is maintained in a "Fixed do" environment. *Music Perception: An Interdisciplinary Journal, 29*(3), 285–296. https://doi.org/10.1525/mp.2012.29.3.285

Wong, Y. K., Lui, K. F. H., Yip, K. H. M., & Wong, A.C.-N. (2020a). Is it impossible to acquire absolute pitch in adulthood? *Attention, Perception, & Psychophysics, 82*(3), 1407–1430. https://doi.org/10.3758/s13414-019-01869-3

Wong, Y. K., Ngan, V. S., Cheung, L. Y., & Wong, A.C.-N. (2020b). Absolute pitch learning in adults speaking non-tonal languages. *Quarterly Journal of Experimental Psychology, 73*(11), 1908–1920. https://doi.org/10.1177/1747021820935776

Open practices statement The data, analysis script, and materials for the study are available on Open Science Framework (https://doi.org/10.17605/OSF.IO/EDAJS). The study was also preregistered through AsPredicted (https://aspredicted.org/n79ey.pdf)

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