Absolute pitch can be learned by some adults

Authors:

Stephen C. Van Hedger[1, 2], Shannon L.M. Heald[1, 2], Howard C. Nusbaum[1, 2]

Affiliations:

[1]Department of Psychology, The University of Chicago: Chicago, IL U.S.A.
[2]Center for Practical Wisdom, The University of Chicago: Chicago, IL U.S.A.

Corresponding Author:
Stephen C. Van Hedger
5848 S. University Avenue
Chicago, IL 60637
svanhedger@uchicago.edu

Word Count (Main Body): 4,933
Tables: 1
Figures: 2

Abstract

Absolute pitch (AP), the rare ability to name any musical note without the aid of a reference note [1–3], is thought to depend on acquisition during an early critical period of development [4–6]. Although recent research has shown that adults can improve AP abilities in a brief (60 min) training session, the best learners in that study still did not achieve note classification performance comparable to performance of a genuine AP possessor [7]. For the first time, we demonstrate that genuine AP levels of performance can be trained in eight weeks for some adults, with the best learner passing all measures of AP ability after training and retaining this knowledge for at least four months after training. The AP measures included the same used in previous assessments of AP ability [8–10]. These results reject strong critical period accounts of AP acquisition, demonstrating clearly that explicit perceptual training in adults can lead to AP performance that is indistinguishable from AP that results from childhood development.

Keywords: *absolute pitch, working memory, perceptual learning, skill acquisition*

167 words

Absolute pitch (AP) is the rare ability to name any musical note without the aid of a reference note [1–3]. The question of how individuals are able to acquire this ability continues to be a matter of scientific debate. The most generally accepted theory is that "genuine" AP ability can only manifest from early-life musical training (the *critical period theory*) [6,11]. In part, the *critical period theory* of AP is bolstered by the lack of conclusive evidence that AP can be learned by post-critical period adults [12–15] as well as a more recent study suggesting that there is a need to re-open the critical period for learning AP with a pharmacological intervention [4].

Here, we directly test the hypothesis of a critical period for AP acquisition through intensive AP training in a post-critical period adult sample. One key difference between this study and previous AP training studies is that participants were selected specifically based on auditory working memory (WM) abilities. Recent research shows that individual differences in auditory WM predict how well adults can learn AP categories from a single training session, although performance was well below genuine AP possessors [7]. Given this observed relationship between auditory WM and AP learning, we asked whether providing substantially more AP training for adults with high auditory WM abilities can produce performance levels comparable to the "genuine" AP listener, specifically with regard to retention of learning (over months) and generalization to novel timbres and tests. This type of demonstration, even in a single adult without prior AP ability, would challenge the *critical period theory* of AP.

As an alternative to the *critical period theory*, AP can be conceptualized as an auditory skill [16] that is shaped by both short- and long-term experiences [17, 18]. This conceptualization predicts that listeners should be able to improve AP performance through explicit perceptual training at any age, with at least some individuals exceeding typical thresholds for genuine AP inclusion post-training. If, however, individuals are only able to modestly improve in AP categorization, with no individual exceeding AP thresholds post-training, this suggests that there might be fundamental limits that constrain performance as would be expected under a critical period framework. Thus, the question of whether a post-critical period adult can learn AP has important implications for the understanding individual differences in AP performance, the underlying mechanisms of AP, and the importance of environmental factors on developing and maintaining AP ability.

Six adult participants completed an eight-week AP training regime (see *Method* for details). Prior to training, participants completed tests of auditory WM, short-term memory (STM), and AP ability. The eight-week AP training program was divided into two phases lasting four weeks each. In both phases of training, participants had to complete three training protocols four times each week, which amounted to approximately 4 hours of training per week (32 hours in total). Every week, participants also completed an AP test in which isolated notes were categorized without feedback. In the second phase of training, we added an additional test that required participants to label the key signature of a presented melody without feedback. At the end of the eight-week training program, participants were retested on the same AP tests that were administered prior to training, in addition to an AP test not previously administered but widely used in prior research. Approximately four months later, we retested participants on the same AP tests as those administered post-training to assess whether learning remained stable.

Results from the pretest session confirmed that all participants scored well on the measures of auditory WM and STM. The average auditory n-back (NB) scores (using d-prime measurements [19]) were 4.05 (*SE*: 0.17) for the 2-back, and 3.52 (*SE*: 0.32) for the 3-back. For reference, in a previous study [7] the average auditory 2-back score was 3.37 (*SE:* 0.16) and the average 3-back score of 2.27 (*SE:* 0.19) from a sample primarily consisting of University of Chicago undergraduates. The average error in the implicit note memory (INM) task was 2.75 (*SE*: 0.30) steps (corresponding to 27.5 cents) meaning that participants were able to recreate a briefly presented sine wave tone within approximately one-quarter of one semitone. For reference, the error of genuine AP possessors performing the identical task is on average 2.44 (*SE:* 0.19) steps [20]. In an auditory digit span (ADS) task, participants correctly recalled an average of 8.83 (*SE*: 0.65) spoken digits, which was between the range of non-AP musical controls (8.1) and genuine AP possessors (10.0) previously reported [21].

Prior to training, none of the participants reached or exceeded the AP cutoffs that have been previously established in the administered external AP Test. [9] We scored this external test (referred to as the UCSF Test), which tested note recognition for both sine wave and piano timbres, by giving full credit for a correct answer and three-quarters credit for answers that fell within one semitone of the correct answer, as this scoring rubric has been previously adopted for this test.[8] For all tests, we also calculated participants' mean absolute deviation (MAD) from the correct note, with scores of 0 reflecting perfect performance and scores of 3 reflecting chance

performance (uniform errors ranging from 0 to 6 semitones removed). Out of a maximum score of 36, prior to training, participants averaged 8.92 points in the Sine Test (MAD: 2.79, *SE*: 0.24) and 10.38 points in the Piano Test (MAD: 2.62, *SE*: 0.29). This level of performance was below the cutoff for the highest designation of AP (*AP*-1), defined as a minimum score of 24.5 on the Sine Test, as well as below the cutoff for the lowest AP designation (*AP-4*), defined as a minimum score of 27.5 on the Piano Test, as specified by the creators of the test.[9] Moreover, all participants scored within the range of non-AP participants from prior investigations using this test [9]. The distribution of pre-training responses for each participant is shown in Figure 1A (Piano) and Figure 1B (Sine), with average MAD represented in Figure 1D. All participant scores (out of 36) are reported in Table 1.

Despite not granting credit for semitone errors on the Chicago AP Test, which consisted of Piano and Multiple Timbre Blocks and was created for the purposes of the present study, we observed higher performance compared to the UCSF Test, likely because the Chicago Test was self-paced. Participants achieved 38.75% accuracy for the Piano Block (MAD: 1.64, *SE*: 0.46) and 34.17% accuracy for the Multiple Timbre Block (MAD: 1.67, *SE*: 0.45). Participants' mean response times (RTs) were 5888ms (*SE*: 1094ms) for the Piano Block and 5199ms (*SE*: 979ms) for the Multiple Timbre Block, which was slower than the fixed presentation rate of notes in the UCSF Test. Based on these pretest scores, our participants either could be classified as non-AP possessors or, at best, "quasi-AP" possessors [22]. The distribution of responses for each participant are represented Figure 2A (Piano Block) and Figure 2B (Multiple Timbre Block), with mean accuracy represented in Figure 2C and mean response time represented in Figure 2D. All participants' accuracy and RTs are additionally reported in Table 1.

By the end of the eight-week training period, participants S2 and S5 were nearly perfect on the Chicago Test. Participant S2 achieved 97.92% accuracy on the Piano Block (MAD: 0.02) and 100% on the Multiple Timbre Block (MAD: 0), while participant S5 achieved 100% on both the Piano Block and Multiple Timbre Block (MAD: 0). For reference, the third highest performing participant scored 52.08% on the Piano Block (MAD: 0.92) and 54.17% on the Multiple Timbre Block (MAD: 0.88), which demonstrates clear above-chance performance (represented by 1/12, or 8.33%) but does not reach thresholds typically used to classify AP possessors.

Within one week after training had ended, we retested all participants inside the lab on the UCSF Test as well as an additional external test of AP ability, referred to as the UCSD Test [10]. Given that these external tests of AP were timed, we interpreted performance on these tests as a more conservative measure of AP learning. Participant S2 showed considerable improvement on the UCSF Piano and Sine Tests, scoring 34.5 on the Piano Test (MAD: 0.05) and 29.5 on the Sine Test (MAD: 0.20) – a level that was above the cutoff for *AP-1* ability. Participant S5 also showed substantial improvements, scoring 24.75 on the Piano Test (MAD: 1.32) and 17.75 on the Sine Test (MAD: 1.89), though participant S5 missed the *AP-1* qualification by 6.75 points and missed the *AP-4* qualification by 2.75 points. All other participants missed the cutoff for *AP-1* by at least 13 points and the cutoff for *AP-4* by at least 10.75 points (Table 1). On the UCSD Test, Participant S2 scored 94.44% (MAD: 0.03), while Participant S5 scored 97.22% (MAD: 0).[1] This level of performance on the UCSD Test qualifies both participants as "AP possessors" based on previous interpretations of this test (using an 85% accuracy cutoff for conservative AP inclusion) [10]. As a comparison, the third highest scoring participant achieved 44.4% accuracy (MAD: 1.24), which once again was above chance by over 36 percentage points but below typical the threshold used to classify AP possessors. The distribution of responses for each participant on the UCSD Test are represented Figure 1C, with average MAD represented in Figure 1D.

To assess the stability of AP category learning, we retested all participants approximately four months (*M* = 128.17 days, *SD* = 6.71 days, range of 117 – 134 days) after training had ended. No participant had reported actively rehearsing pitch-label associations in the time between the AP posttests and the follow-up tests. Previous AP training research has been criticized for not following up with participants after training has ended to see how category learning is retained [23], as genuine AP possessors appear to have relatively stable categories that do not require explicit maintenance (though see [17,18,24] for alternate views). We administered the same AP assessments given to participants in the posttest (UCSF Test, the UCSD Test, and the Chicago Test).

Results from the follow-up AP tests supported the conclusion that Participants S2 and S5 retained stable performance across all tests of AP ability. For the UCSF Test, Participant S2 still

[1] The discrepancy between accuracy and MAD for participant S5 was because one note was not labeled at all. As such, it was counted as incorrect but could not be used in the calculation of MAD.

passed the *AP-1* cutoff, scoring 32.75 on the Sine Test and 30.75 on the Piano Test. Participant S5 retained high performance (13.75 on Sine, 24.75 on Piano), but was still below the *AP-1* cutoff by 10.5 points and missed the *AP-4* designation by 2.75 points. For the UCSD Test, Participant S2 scored 77.78% (MAD: 0.22) and Participant S5 scored 91.67% (MAD: 0.31). Even though Participant S2 fell below the 85% cutoff used as a conservative inclusion measure in prior studies, S2 never missed by more than one semitone. Adopting the more liberal inclusion measures from prior studies that have used the UCSD Test (in which semitone errors are allowed), Participant S2 would still be categorized as possessing AP [10]. It should be noted that Participant S2 also took the UCSD Test again approximately 16 months post-training for a separate study and achieved 88.89% accuracy (MAD: 0.14) not including semitone errors as correct. Finally, for the Chicago Test, Participants S2 and S5 scored identically in terms of percentage correct (93.75% on the Piano Block, MADs of 0.06 and 0.10, respectively) and 91.67% on the Multiple Timbre Block, MADs of 0.11 and 0.13, respectively).

These results challenge two commonly held theoretic assumptions regarding AP – (1) that AP is an “all or none” ability [8] and (2) that “genuine” AP cannot be trained in adults [1]. While one could argue that Participants S2 and S5 were always AP possessors, and thus our training program merely revealed an inherent ability, we find this explanation unlikely because of both participants’ previous musical instruction. Specifically, both Participants S2 and S5 had extensive musical training (beginning musical instruction at the ages of 7 and 6, respectively, and playing their primary instrument for at 8 and 20 years, respectively). Thus, both participants had considerable previous experience with pitch-label mappings, yet did not exceed conventional performance thresholds for classifying AP ability prior to training. However, it is clear that both Participants S2 and S5 had at least some absolute pitch knowledge prior to training, as they scored above all other participants on all pretests of AP. Thus, the present results cannot speak to whether anyone, given enough training, could develop sufficient speed and accuracy in note labeling as to be considered an AP possessor. Indeed, it is possible that participants S2 and S5 may have possessed a kind of latent AP knowledge from their musical experiences that was refined through perceptual training in the present study. As such, we cannot entirely refute a critical period account of AP, even if we offer the most comprehensive evidence to date that adult AP acquisition is possible in some individuals.

One might also argue that what we observed is not genuine AP acquisition. If this argument is to be made, however, then the current definition of AP needs to be reconsidered, as both Participants S2 and S5 performed at levels equivalent to genuine AP possessors on tests of AP ability, and both participants retained AP performance several months after training had ended.

While the present study is limited in sample size, the demonstration that even two adults can, with moderate training, reach typical "true" AP levels of performance is theoretically important given the strong claims made about the dependence of AP on critical-period learning. No prior published study has demonstrated a comparable level of successful adult AP learning and long-term retention, measured through previously established, widely adopted, and strictly timed tests. As such, we stress the importance of these findings as proof-of-concept that adult AP acquisition is possible and that learning remains stable months after explicit training has ended. To this end, *any* demonstration of successful AP learning by an adult will inform the discussion of the underlying mechanisms of AP acquisition and maintenance. It is also important to consider that these levels of AP performance were achieved with eight weeks (approximately 32 hours) of adult training, which is far less than the amount of training that is typically thought to be required for the explicit learning of AP in childhood based on prior research [25].

Importantly, the present results can be situated in a broader context treating AP as an auditory skill. The current study is the most conclusive demonstration that AP can be acquired and retained by post-critical period adults with exceptional auditory memories. These findings resonate with the growing body of work demonstrating that auditory WM abilities relate to (1) pitch memory accuracy for familiar musical recordings [26], (2) the rapid and explicit training of AP among previously naïve individuals [7], and (3) note categorization accuracy among genuine AP listeners under challenging listening conditions [27]. In conclusion, the present results challenge the theoretical claim that AP must manifest within a critical period of development and refocuses future inquiry regarding AP on understanding the cognitive mechanisms that mediate AP acquisition.

Method

Participants

Six participants (three female, three male) participated in the experiment (*M* = 23.33, *SD* = 2.94 years old, age range: 18-26). None of the participants self-reported possessing AP. All participants provided written consent and were compensated for their participation in the experiment.

Task Description and Materials

AP Training Program

The absolute pitch training program was eight weeks in duration. There were two phases to the experiment (each lasting four weeks). Both phases consisted of three training programs that participants had to complete four days every week. Participants were tested every Friday, meaning that participants could complete their weekly training programs Saturday through Thursday.

First Phase. Two of the First Phase training programs were meant to emphasize speed in classifying absolute pitches, while the third was meant to emphasize accuracy. In the first program, nicknamed “Simple Speed” (SS), participants would see a note name presented in the center of the screen for 1500ms (e.g., C). This was the target note for the trial. During the presentation of the note name, participants heard a sung version of the target note, which was enunciated with the category label. After participants heard the sung target note, they heard a string of 16 notes, with an inter-note-interval of 2250ms. Within these 16 notes, 25% (4 of 16) were the target note, while 75% (12 of 16) were non-target notes. Target notes were randomly interspersed with non-target notes. Participants had to press the spacebar as quickly as possible whenever they heard the target note. Participants had 1750ms to respond to target notes. All notes were taken from a C-major scale (i.e., white keys only) spanning a one-octave range (C [4] to B [4]). Non-target notes were also taken from a one-octave range and consisted of only white keys. The notes were synthesized with a piano timbre. Participants received feedback after each note. There was a 1500ms rest period between trials. Each of the seven target notes (C, D, E, F, G, A, and B) were presented twice, for a total of 14 trials during each session.

The second program, nicknamed “Complex Speed” (CS), followed the same general procedure as SS, with the following differences. First, the presentation of the 16 notes was faster (1750ms), meaning that participants had to respond more quickly (within 1250ms). Second, the notes (both target and distractor) could come a piano, flute, or guitar timbre. Third, the octave range was expanded relative to the SS task (C [3] to B [5]).

The third program, nicknamed “Accuracy Training” (AT), emphasized correct note category identification over speed of identification. There were two blocks, consisting of 48 trials each (96 trials total). The general procedure was the same for both blocks. On each trial, participants would hear an isolated note, which was presented simultaneously with a response screen. The response screen displayed all twelve note-category options, arranged in a pitch wheel. Participants then had an unlimited amount of time to click on the corresponding note name with the mouse. After each response, participants would receive feedback, in which the correct note name was highlighted and participants reheard the note. The first block consisted of piano notes, spanning a two-octave range (C [4] to B [5]). Unlike the SS and CS tasks, all 12 categories were explicitly trained. Since there were 24 total piano notes (12 note categories x 2 octaves), each note was presented twice in a random order. The 48 notes in the second block were randomly selected from 96 possible notes (24 piano, 24 cello, 24 clarinet, and 24 harpsichord, see Table S1). Participants heard 1000ms of noise between trials to minimize the influence of relative pitch information in selecting the correct absolute pitch.

Every Friday, participants would have to complete a weekly test (WT) of AP ability. The WT was similar to the AT task, with two exceptions. First, participants did not receive feedback. Second, the second block was expanded to 120 possible notes (24 piano, 24 cello, 24 clarinet, 24 harpsichord, and 24 square wave).

Second Phase. We replaced the SS and CS tasks after Week 4 with a more difficult speeded task, nicknamed “Hypercomplex Speed” (HS). The HS task followed the same general procedure as the SS and CS, with the following exceptions. First, on each trial, participants heard a string of 32 notes, with an inter-note-interval of 1500ms. Within these 32 notes, 12.5% (4 of 32) were the target note, while 87.5% (28 of 32) were non-target notes. Second, the response decision window was reduced to 1000ms. Third, both the target and the non-target notes were taken from all 12 note categories (sampling from all 160 note stimuli described in Table S1).

We also introduced a novel task during the Second Phase nicknamed “Name That Key” (NTK). The NTK task followed the same general procedure as the AT task. Each session, participants heard 15 music recording excerpts, which were randomly selected from a database of 300 total recordings. Similar to the AT task, participants had to decide on the key signature of the recording by clicking on a pitch wheel with the 12 pitch categories arranged around a circle. After each selection, participants would see a specific feedback screen, in which the correct note name was highlighted on the pitch wheel. Additionally, participants heard the tonic note of the key signature (e.g., C) played during feedback. Participants heard 1000ms of noise between trials. Participants also completed the AT task during the Second Phase, which was identical to the version presented in the First Phase.

Every Friday, participants would have to complete the same WT from the First Phase, in addition to a new test, nicknamed the “Name That Key Test” (NTKT). The NTKT required participants to judge the key signature of 12 folk songs. During the task, each pitch class was represented by a key signature one time (i.e., one folk song played in C, one folk song played in C#, etc.), and feedback was not provided. The randomized assignment of folk melody to key signature was hard coded into the script, as we did not want any given folk song to play in the same key signature across weeks.

The isolated musical note stimuli used in training were created using Reason Music Production Software (Propellerhead: Stockholm, Sweden). Musical notes were sampled from real instruments and were digitized at a 44.1 kHz sampling rate with 16-bit depth, were 1000ms in duration, and were root mean square normalized to a level of -5 dB FS. We used a total of 160 notes (from seven instrumental timbres) throughout training. Details of the specific notes used can be found in Table S1. For the music recording stimuli used in the NKT task, we recorded 15-second excerpts from 300 popular pieces of music (e.g., pop songs, movie themes). For the NTKT task, we recorded simple piano melodies of folk songs using Reason. We then transposed and exported each folk melody in every key. The explicit absolute pitch training and testing programs were run on participants’ personal computers using Open Sesame software [28]. All stimuli and training scripts are available on the Open Science Framework.

Tests of Absolute Pitch

We assessed AP ability before and after the training program using several tests of AP. The UCSF AP Test has been used in several prior studies of AP [8,9], and could be accessed through the University of California San Francisco AP website. We recorded the audio stimuli (40 piano tones, 40 sine wave tones) from the website and administered the test offline and in the lab to monitor the participants (minimizing the possibility of using non-absolute cues or strategies such as humming to artificially inflate their scores). We removed the four highest tones for the sine wave test and the four lowest tones for the piano test, which is standard for scoring the UCSF Tests. Participants identified each note in writing.

Both the UCSF Piano and Sine Tests consisted of 40 trials. Each tone lasted 1000ms with approximately a 2250ms interlude between tone onsets (meaning the inter-note-interval was approximately 3250ms). There was a longer break (10s) after every 10 notes, which gave participants the opportunity to check to make sure that they were categorizing the correct trial number. Under no circumstances were participants allowed to repeat any of the notes. Participants were randomly assigned to take either the Piano or Sine Test first.

The UCSD Test was taken from Diana Deutsch's website (deutsch.ucsd.edu), and has been used to assess AP ability across a wide variety of participants [10]. To minimize the use of relative pitch as a cue, all intervals between successive notes were larger than an octave. We downloaded the audio file from the website and administered the test offline and in the lab to monitor the participants (minimizing the possibility of using non-absolute cues or strategies such as humming to artificially inflate their scores). The participants identified each note in writing.

The UCSD Test consisted of 36 scored piano notes, which spanned from C [3] (below middle C) to B [5] (almost three octaves above middle C) divided into three block of 12 notes. The first four notes were not scored and were designated as practice trials, which is standard for this test. Each piano note lasted approximately 500ms, and there was approximately 3750ms of silence between notes (meaning the inter-note-interval was approximately 4250ms). After each block of 12 notes, there was an extended break of approximately 18s, which gave participants the opportunity to check to make sure that they were categorizing the correct trial number. Under no circumstances were participants allowed to repeat any of the notes.

The Chicago Test was identical to the WT from training. We treated the Piano Block separately from the Multiple Timbre Block.

Tests of Auditory Working Memory

The implicit pitch memory (INM) task has been previously used as a test of auditory working memory precision[7], and has also been associated with both explicit and implicit AP representations[7,26]. On each trial, participants heard a brief (200ms) sine wave target note, which was then masked by 1000ms of noise. Participants then had to adjust a starting note, by clicking on upward and downward arrows on the computer screen, to try to recreate the originally heard target note. The arrows moved the pitch either 10 or 20 cents up or down, depending on whether participants were clicking on the smaller arrows (10 cents) or larger arrows (20 cents). 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 27 sine waves in the distribution. The lowest frequency was 471.58 Hz, corresponding to a 20-cent sharp Bb [4] and the highest frequency was 547.99 Hz, corresponding to a 20-cent flat C# [5]. The intermediary frequencies were evenly spaced in 10-cent increments. On each trial, participants had to recreate one of 10 possible targets (corresponding to pitches 5, 7, 9, 11, 13, 15, 17, 19, 21, or 23 in the distribution), starting from one of four fixed locations (corresponding to pitches 1, 3, 25, or 27 in the distribution). Participants randomly heard all combinations of target / starting location twice, resulting in 80 trials (10 target notes x 4 starting notes x 2 repetitions). The INM task was run using the Psychophysics Toolbox in Matlab. [29,30]

The auditory n-back (NB) task required participant to monitor the identity of spoken letter strings and remember whether the current spoken letter matched the letter presented *n* trials previously. All participants completed a 2-back and a 3-back task. For all the trials in which the current letter did not match the spoken letter presented n trials previously, participants were instructed to press a button labeled “Not Target.” Both the 2-back and 3-back consisted of 90 trials (three runs of 30 spoken letters). Letters were spoken sequentially, with an inter-stimulus-interval of 3000ms. Targets occurred one-third of the time, while non-targets occurred two-thirds of the time. Before the 2-back and 3-back, participants completed a practice round of 30 trials to familiarize themselves with the task. The NB was run in E-Prime 2.0 (Psychology Software Tools: Sharpsburg, PA).

We assessed auditory short-term memory (STM) using the auditory digit span (ADS) task. For the ADS task, participants initially heard three trials of five-number strings (e.g.,

27483). There was 1000ms of silence between spoken numbers. Participants needed to correctly identify a majority (at least two) of the five-number strings in order to advance to six-number strings. If participants could not correctly identify a majority of the five-number strings (attaining zero or one correct answer), they were given three trials of four-number strings. This process of adding or removing a number based on performance repeated eight times. Thus, a perfect performance would yield a digit span score of 13 (5+8), while a completely inaccurate performance would yield a digit span score of 1. The ADS was run in E-Prime 2.0.

Questionnaires

Participants filled out a music experience questionnaire at the completion of training, which asked about primary music experience, the number of instruments played (as well as the number of years of active musical instruction on each instrument), and the age of beginning musical instruction. Additionally, participants filled out a follow-up questionnaire, which specifically asked about participants' explicit AP training in the time between posttest and follow-up.

Procedure

After providing informed written consent, participants completed the NB, INM, and ADS measures, which were followed by the AP Tests (UCSF and Chicago). These were completed in the lab over in a single session. Participants were then given flash drives containing the AP Training Program, which contained detailed instructions for completion. Participants were additionally walked through the general training protocol by the experimenter.

Over the next eight weeks, participants completed both phases of their training program as specified by the instructions. One participant was unable to complete three days of training in Week 7. Every Friday, participants completed their weekly tests. Participants uploaded their data to a secure server after each week.

Within one week of completing the eight-week training program, participants returned to the lab to complete the external AP tests (UCSF and UCSD), in addition to a musical experience questionnaire. We treated participants' final WT as their Chicago posttest, as this was completed on the final day of the training program. No feedback was given. The follow-up test, which was identical in design to the posttest, was conducted approximately four months after training had

ended depending on availability of the participants. Around the same time as the follow-up test, participants completed an online questionnaire that assessed their musical activities in the period between the posttest and follow-up, including explicit AP practice.

Acknowledgements

This research was supported by the Multidisciplinary University Research Initiatives (MURI) Program of the Office of Naval Research through grant, DOD/ONR N00014-13-1-0205.

Note on Open Materials

All data and training programs associated with this paper are available on the Open Science Framework (https://osf.io/9n48c/)

References

1. Takeuchi, A. H. & Hulse, S. H. Absolute pitch. *Psychol. Bull.* **113,** 345–361 (1993).
2. Deutsch, D. Absolute Pitch. in *The Psychology of Music* (ed. Deutsch, D.) 141–182 (Academic Press, 2013). doi:10.1016/B978-0-12-381460-9.00005-5
3. Ward, W. D. & Burns, E. M. Absolute pitch. in *The Psychology of Music* (ed. Deutsch, D.) 431–451 (Academic Press, 1982).
4. Gervain, J. *et al.* Valproate reopens critical-period learning of absolute pitch. *Front. Syst. Neurosci.* **7,** 1–11 (2013).
5. Russo, F. A., Windell, D. L. & Cuddy, L. L. Learning the 'special note': Evidence for a critical period for absolute pitch acquisition. **21,** 119–127 (2003).
6. Levitin, D. J. & Zatorre, R. J. On the Nature of Early Music Training and Absolute Pitch: A Reply to Brown, Sachs, Cammuso, and Folstein. *Music Percept.* **21,** 105–110 (2003).
7. Van Hedger, S. C., Heald, S. L. M., Koch, R. & Nusbaum, H. C. Auditory working memory predicts individual differences in absolute pitch learning. *Cognition* **140,** 95–110 (2015).
8. Athos, E. A. *et al.* Dichotomy and perceptual distortions in absolute pitch ability. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 14795–14800 (2007).
9. Baharloo, S., Johnston, P. a, Service, S. K., Gitschier, J. & Freimer, N. B. Absolute pitch: an approach for identification of genetic and nongenetic components. *Am. J. Hum. Genet.* **62,** 224–231 (1998).
10. Deutsch, D., Henthorn, T., Marvin, E. & Xu, H. Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.* **119,** 719–722 (2006).
11. Crozier, J. B. Absolute pitch: Practice makes perfect, the earlier the better. *Psychol. Music* **25,** 110–119 (1997).
12. Hartman, E. B. The influence of practice and pitch-distance between tones on the absolute identification of pitch. *Am. J. Psychol.* **67,** 1–14 (1954).
13. Cuddy, L. L. Practice effects in the absolute judgment of pitch. *J. Acoust. Soc. Am.* **43,** 1069–1076 (1968).
14. Lundin, R. W. Can perfect pitch be learned? *Music Educ. J.* **49,** 49–51 (1963).
15. Vianello, M. a & Evans, S. H. Note on pitch discrimination learning. *Percept. Mot. Skills* **26,** 576 (1968).
16. Bermudez, P. & Zatorre, R. J. A distribution of absolute pitch ability as revealed by computerized testing. *Music Percept.* **27,** 89–101 (2009).
17. Hedger, S. C., Heald, S. L. M. & Nusbaum, H. C. Absolute Pitch May Not Be So Absolute. *Psychol. Sci.* **24,** 1496–1502 (2013).
18. Wilson, S. J., Lusher, D., Martin, C. L., Rayner, G. & McLachlan, N. Intersecting factors

lead to absolute pitch acquisition that is maintained in a ‘Fixed do’ Environment. *Music Percept.* **29,** 285–296 (2012).

19. Macmillan, N. A. & Creelman, C. D. *Detection theory: A user’s guide*. (Lawrence Erlbaum Associates, 2005).

20. Heald, S. L. M., Van Hedger, S. C. & Nusbaum, H. C. Auditory category knowledge in experts and novices. *Front. Neurosci*. **8,** 1–15 (2014).

21. Deutsch, D. & Dooley, K. Absolute pitch is associated with a large auditory digit span: A clue to its genesis. *J. Acoust. Soc. Am.* **133,** 1859–1861 (2013).

22. Bachem, A. Various types of absolute pitch. *J. Acoust. Soc. Am.* **9,** 146–151 (1937).

23. Rush, M. An experimental invesitagtion of the effectiveness of training on absolute pitch in musicians. (The Ohio State University, 1989).

24. Van Hedger, S. C., Heald, S. L. M., Uddin, S. & Nusbaum, H. C. A note by any other name: Intonation context rapidly changes absolute note judgments. *J. Exp. Psychol. Hum. Percept. Perform.* (2018). doi:10.1037/xhp0000536

25. Miyazaki, K. & Ogawa, Y. Learning Absolute Pitch by Children. *Music Percept.* **24,** 63–78 (2006).

26. Van Hedger, S. C., Heald, S. L. M. & Nusbaum, H. C. Long-term pitch memory for music recordings is related to auditory working memory precision. *Q. J. Exp. Psychol.* 1–13 (2017). doi:10.1080/17470218.2017.1307427

27. Van Hedger, S. C. & Nusbaum, H. C. Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background. *PsyArXiv* (2018). doi:http://doi.org/10.17605/OSF.IO/VFQE6

28. Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* **44,** 314–324 (2012).

29. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10,** 433–436 (1997).

30. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial vision* **10,** 437–442 (1997).

Tables

Table 1: Performance measures across the Chicago Piano (C-P) and Multiple Timbre (C-MT) Blocks, the UCSF Piano (UCSF-P) and Sine (UCSF-S) Tests, as well as the UCSD Test

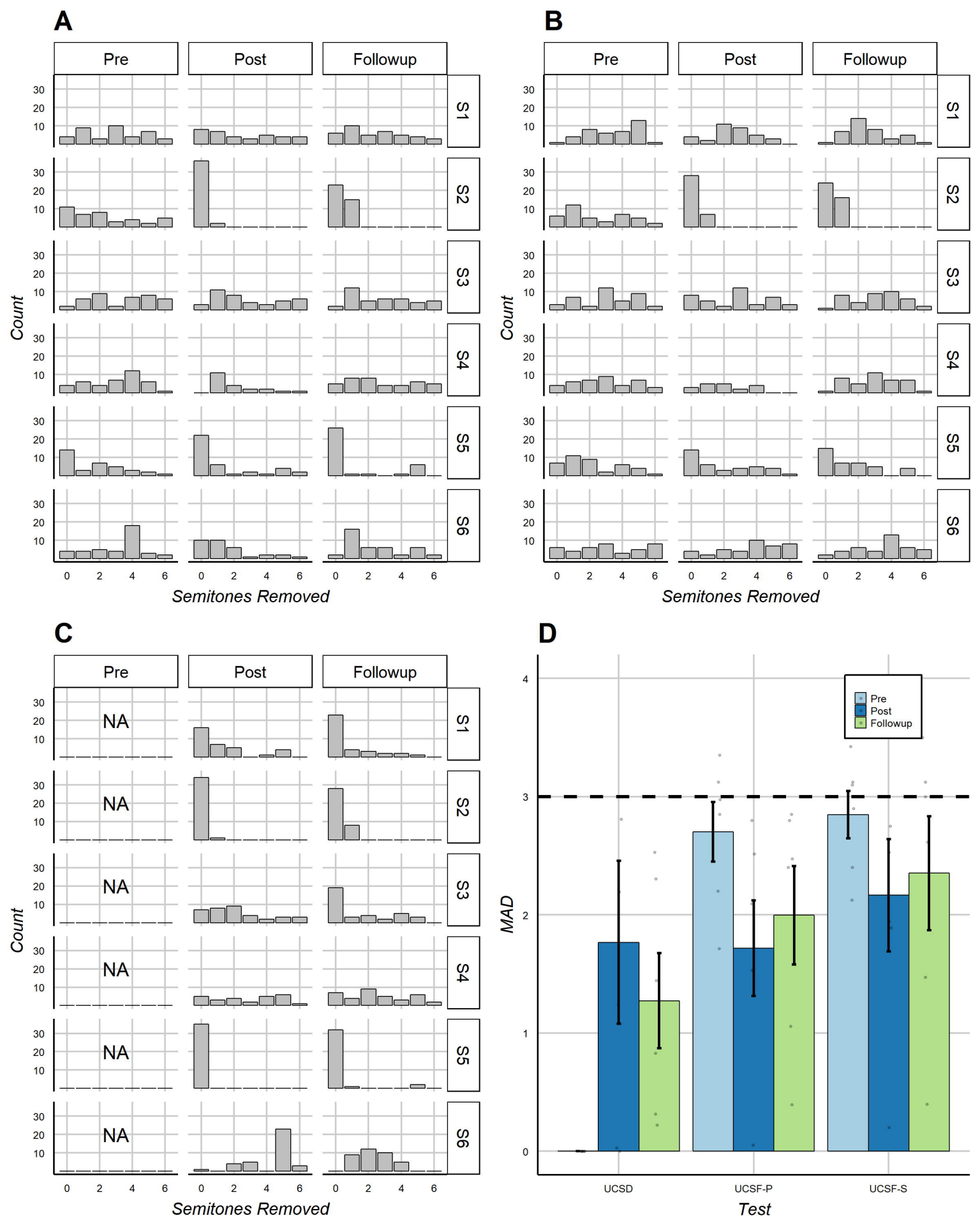
Pretest Participant	C-P	C-MT	UCSF-P	UCSF-S	UCSD
S1	25.00% (8.16s)	25.00% (7.01s)	9.25	3	*NA*
S2	62.50% (3.66s)	52.08% (3.22s)	16.25	12.75	*NA*
S3	14.58% (4.52s)	12.50% (4.25s)	5.75	7.5	*NA*
S4	20.83% (8.90s)	12.50% (8.04s)	8.5	8.5	*NA*
S5	70.83% (4.19s)	68.75% (3.48s)	15.5	13.5	*NA*
S6	6.25% (7.75s)	4.17% (7.53s)	7	8.25	*NA*
Posttest Participant	C-P	C-MT	UCSF-P	UCSF-S	UCSD
S1	52.08% (7.08s)	54.17% (6.25s)	13.25	4.75	41.67%
S2	97.92% (1.59s)	100% (1.57s)	34.5	29.5	94.44%
S3	37.50% (2.94s)	31.25% (2.41s)	10.5	11.5	19.44%
S4	47.92% (4.97s)	52.08% (5.12s)	7.5	6	11.11%
S5	100% (3.07s)	100% (2.88s)	24.75	17.75	97.22%
S6	0% (2.38s)	0% (2.81s)	16.75	5.5	2.78%
Follow-up Participant	C-P	C-MT	UCSF-P	UCSF-S	UCSD
S1	56.25% (6.48s)	52.08% (5.84s)	13.5	6.25	63.89%
S2	93.75% (1.77s)	91.67% (1.80s)	30.75	32.75	77.78%
S3	33.33% (3.04s)	47.92% (2.63s)	9.5	7	52.78%
S4	18.75% (6.81s)	22.92% (6.06s)	11	7	19.44%
S5	93.75% (3.46s)	91.67% (2.95s)	24.75	18.25	91.67%
S6	2.08% (3.84s)	0% (3.23s)	13.25	5	0%

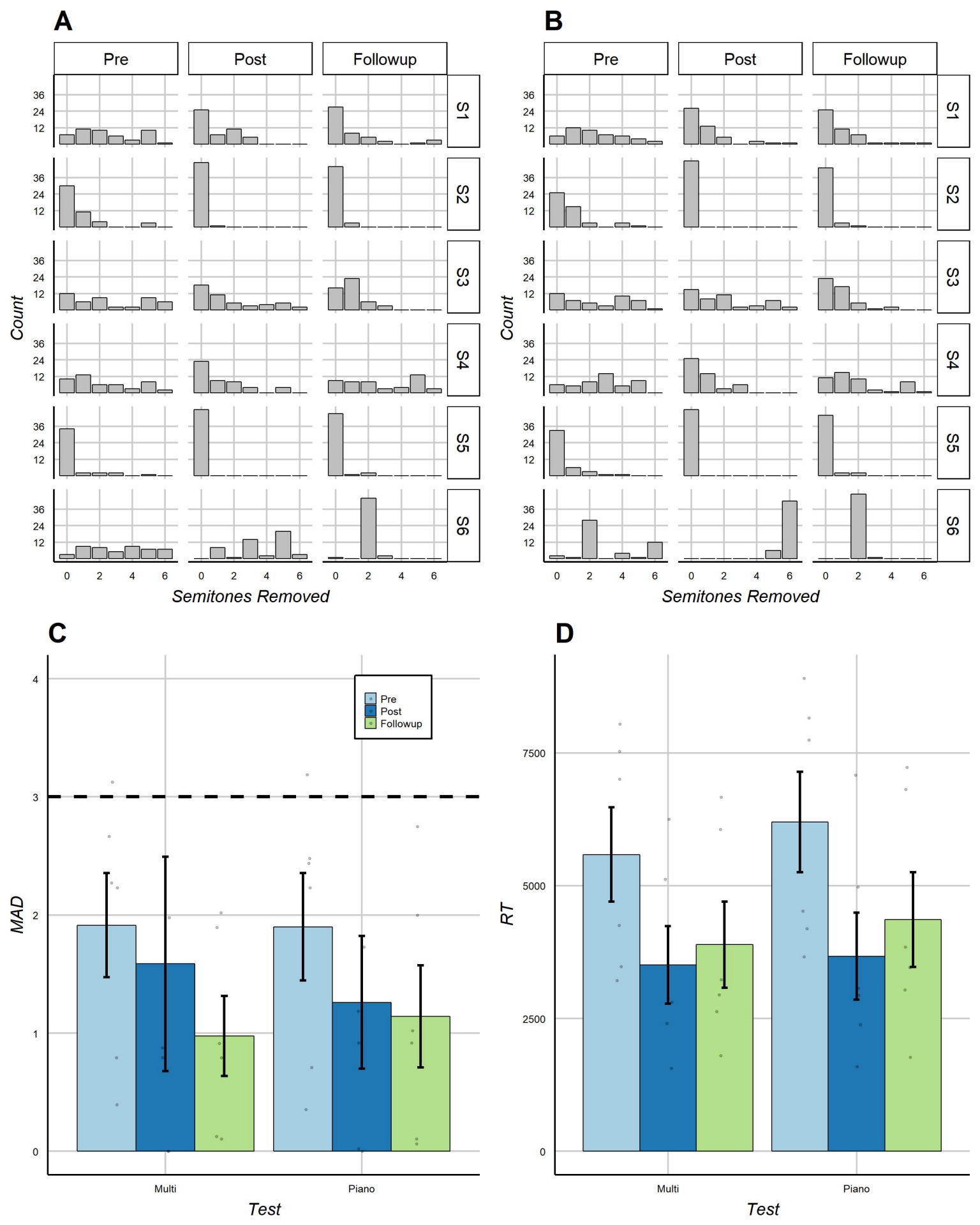
Note: Parentheses following the SD percentages represent mean response time. UCSF values represent performance (out of 36 scored trials), where full credit was given for correct answers and three-quarters credit was given for incorrect answers within one semitone of the correct answer. UCSD values represent percentage correct (no credit for semitone errors).

Figure Captions

Figure 1: Distribution of semitone errors for each participant (S1-S6) and each testing session (Pre, Post, and 4-month delayed Follow-up) for the UCSF Piano Test (A), UCSF Sine Test (B), UCSD Test (C), and mean absolute deviation across all participants for these tests (D).

Figure 2: Distribution of semitone errors for each participant (S1-S6) and each testing session (Pre, Post, and 4-month Follow-up) for the Note ID Test, split by the Piano Block (B) and Multiple Timbre Block (B). Mean absolute deviation across all participants for this test is represented in panel C, whereas mean response time across all participants for this test is represented in panel D.


A
B
C
D
Pre
Post
Followup
S1
S2
S3
S4
S5
S6
Count
Semitones Removed
NA
MAD
Test
UCSD
UCSF-P
UCSF-S


A
B
Pre
Post
Followup
S1
S2
S3
S4
S5
S6
Count
Semitones Removed
C
D
Pre
Post
Followup
MAD
RT
Multi
Piano
Test