

RESEARCH ARTICLE

Revisiting discrete versus continuous models of human behavior: The case of absolute pitch

Stephen C. Van Hedger^{1,2,3,4}*, John Veillette^{1,2}*, Shannon L. M. Heald^{1,2}, Howard C. Nusbaum^{1,2}

1 Center for Practical Wisdom, The University of Chicago, Chicago, IL, United States of America, **2** Department of Psychology, The University of Chicago, Chicago, IL, United States of America, **3** Brain and Mind Institute, Western University, London, ON, Canada, **4** Department of Psychology, Huron University College, London, ON, Canada

* These authors contributed equally to this work.

* svanhedg@uwo.ca (SCVH); johnv@uchicago.edu (JV)

OPEN ACCESS

Citation: Van Hedger SC, Veillette J, Heald SLM, Nusbaum HC (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>

Editor: Lutz Jäncke, University of Zurich, SWITZERLAND

Received: May 4, 2020

Accepted: December 7, 2020

Published: December 28, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0244308>

Copyright: © 2020 Van Hedger et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data and materials can be accessed through Open Science Framework: <https://osf.io/v46aj/>.

Funding: This research was supported in part from a grant from the John Templeton Foundation

Abstract

Many human behaviors are discussed in terms of discrete categories. Quantizing behavior in this fashion may provide important traction for understanding the complexities of human experience, but it also may bias understanding of phenomena and associated mechanisms. One example of this is absolute pitch (AP), which is often treated as a discrete trait that is either present or absent (i.e., with easily identifiable near-perfect “genuine” AP possessors and at-chance non-AP possessors) despite emerging evidence that pitch-labeling ability is not all-or-nothing. We used a large-scale online assessment to test the discrete model of AP, specifically by measuring how intermediate performers related to the typically defined “non-AP” and “genuine AP” populations. Consistent with prior research, individuals who performed at-chance (non-AP) reported beginning musical instruction much later than the near-perfect AP participants, and the highest performers were more likely to speak a tonal language than were the lowest performers (though this effect was not as statistically robust as one would expect from prior research). Critically, however, these developmental factors did *not* differentiate the near-perfect AP performers from the intermediate AP performers. Gaussian mixture modeling supported the existence of two performance distributions—the first distribution encompassed both the intermediate and near-perfect AP possessors, whereas the second distribution encompassed only the at-chance participants. Overall, these results provide support for conceptualizing intermediate levels of pitch-labeling ability along the same continuum as genuine AP-level pitch labeling ability—in other words, a continuous distribution of AP skill among all above-chance performers rather than discrete categories of ability. Expanding the inclusion criteria for AP makes it possible to test hypotheses about the mechanisms that underlie this ability and relate this ability to more general cognitive mechanisms involved in other abilities.

1. Introduction

In psychological science, human behaviors are sometimes quantized to a small number of discrete categories, often reflected as the simple presence or absence of an ability. Often this is

(58345) to HCN (<https://www.templeton.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

done merely for the sake of empirical tractability, as in the cases of executive function [1], musicality [2] or empathy [3]. In other cases, however, human behavior is discretely categorized by researchers because the boundaries between categories are thought to reflect fundamental differences in the underlying processes and mechanisms that mediate the behavior—an assumption that sharply constrains the types of explanations one can apply to understand such behaviors. A particularly salient example of this is absolute pitch (AP), typically defined as the rare ability to name or produce any musical note without a reference note [4,5].

AP has been described an excellent model system for understanding the interaction of genetic and environmental factors, in part because it is thought to represent a discrete ability that individuals either clearly do or do not possess, which makes the potential genetic and environmental markers theoretically easier to isolate [6]. In fact, the observed dichotomy in note labeling ability in some tests (clusters of near-perfect versus at-chance performers with very few intermediate data points) has led researchers to speculate that AP may be governed by just a couple of genes modulated by early auditory experiences occurring within a critical period [7]. This dichotomized, discrete view of AP suggests that there are two kinds of listeners, clearly differentiated by their explicit abilities to categorize auditory pitch based on chroma. The first kind of listener who, upon hearing an isolated musical note, is essentially guessing its note name, can be thought of as a “non-AP possessor”. The second kind of listener who, upon hearing an isolated musical note, can quickly and nearly perfectly categorize the note with its respective note name (e.g., C), can be thought of as a “genuine-AP possessor”.

In this vein, describing potential factors that are responsible for the putative dichotomy in note labeling ability has been part and parcel of AP research for the better part of a century [8–10]. One of the most consistently linked experiential factors differentiating AP and non-AP listeners is musical training within a *critical period* of development [11]. Researchers have robustly found that age of musical training onset predicts whether an individual exceeds the experimenter-defined thresholds of performance on pitch labeling tasks that define “genuine” AP ability [8,10–13]. A second experiential factor that has been associated with AP is tone language experience [14]. Several direct assessments of music students enrolled in top music conservatories in the United States and China have found that tone language speakers perform significantly better on AP tests compared to non-tone language speakers and are more likely to exceed typical thresholds for defining “genuine” AP ability [15,16]. While lexical tone distinctions do not depend on absolute pitches (e.g., the “stable” pitch of Tone 1 in Mandarin does not need to be uttered at a particular absolute frequency), the fact that speakers of tone languages show minimal pitch variability in productions—even across days—suggests that absolute pitch cues may be reinforced at least within each speaker of a tone language [14].

Critically, many of the findings concerning experiential contributions to the development of AP depend on a discrete operationalization of AP in their study recruitment design (i.e. two groups of near-ceiling vs. presumably-at-chance performers) and analyses (discrete vs. continuous statistics). Biological phenotypes, however, are often continuous, not discrete, reflecting complex and continuous interactions of genetic and environmental factors that can contribute to the manifestation of a particular trait (e.g., height or weight). When AP is formalized as a discrete ability [7], as much of the literature at least implicitly assumes, direct comparisons of near-perfect AP possessors and at-chance non-AP possessors are appropriate. If, however, AP is a more continuously distributed ability, then such classifications of individuals into “AP” and “non-AP” groups may remove important variance in absolute pitch-labeling ability that may be vital in fully understanding how these experiences relate to the ability.

There are several reasons to suspect that AP is a more continuously distributed ability than previously believed. First, even among self-identified “AP possessors,” the distribution of pitch-labeling performance depends on the nature of the test materials, such as familiar versus

unfamiliar timbres [17]. Second, operationalizations of AP that are more graded (e.g., assessing musical note labeling beyond the simple correct/incorrect binary) have shown large amounts of performance variability in both self-identified AP and non-AP populations [18]. Both points highlight potential problems in constructing “AP” and “non-AP” groups for comparison, as the threshold for inclusion in an AP group may be arbitrary and may not capture a significant portion of the population with intermediate pitch naming abilities. While there has been research to suggest that AP may indeed be best described as a continuously distributed ability [5,19], it is presently unclear how particular environmental mechanistic explanations, such as critical-period musical training or tone language experience, relate to *gradations* of absolute pitch-labeling ability. One likely reason why this question has received relatively little attention is because intermediate levels of AP performance (sometimes denigrated as “pseudo” AP) are often treated as an entirely separate phenomenon [20]. Yet, performance thresholds for differentiating “genuine” from “pseudo” AP may not be clear and so these designations may be more reflective of a theoretical framework treating AP as categorically discrete rather than a true separation of ability. Given these considerations, the present large-scale, online study had two primary aims. The first aim was to assess independently whether intermediate AP should indeed be treated as a separate phenomenon from “genuine” AP rather than existing along the same AP continuum. The second aim was to assess how experiential factors (specifically, critical-period musical training and tone language experience) relate to AP performance across different levels of AP proficiency.

The present study uses an online assessment of absolute pitch ability to assess the distribution of absolute pitch labeling ability in the general population. We first apply pitch-labeling performance cut-offs similar to those used in past literature to sort the subjects into “genuine-AP,” “pseudo-AP,” and “non-AP” groups, and we assess the distribution of experiential factors that have been commonly related to AP (age of music onset and tonal language experience) across these groups. We then apply Gaussian Mixture Modelling to characterize the distribution of pitch labeling performance in a more data-driven manner, identifying two distinct distributions, and we contrast the natural break between these distributions with the normative cut-offs for AP-level. Throughout, we assess whether common findings from past AP research, in which AP performance is predicted by music and tonal language experience, can be explained distribution membership versus continuous position within each distribution, revealing the extent to which the historically dichotomized conception of AP can bias analytic results.

2. Method

2.1 Participants

We received a total of 195 unique responses (assessed by IP address) to the online AP assessment. These responses spanned a period of approximately 29 months, from May 6, 2016 to October 2, 2018. If multiple responses were associated with a single IP address, we selected the earliest response from each IP to isolate a single respondent. While this culling may appear overly conservative, it should be noted that virtually all the duplicate responses (92.9%) were associated with high levels of note identification accuracy (>80%), suggesting that individuals who performed well were more likely to repeat the measure. Approximately half of the 195 participants (48.6%) completed the study within a one-week period (from June 11, 2017 to June 16, 2017). We attribute this spike in participation to a Wall Street Journal article published on June 11, 2017 [21] that provided a link to the online study.

We culled the data based on three additional considerations prior to analysis. First, we removed participants if they did not complete the brief end-of-study questionnaire, as these

questions were essential for informing our research questions related to age of beginning musical training and tone language background. This consideration removed 25 participants. Second, we removed participants who “timed out” (i.e., did not provide a response) on more than 50% of the trials. This consideration removed an additional 11 participants. Third, we removed participants who reported no musical instruction, as it would not be appropriate to calculate an age of music onset for these individuals. This consideration removed an additional 7 participants. In total, then, 152 participants were included in the analyses. All individuals who completed the online study provided informed consent and the protocol was approved by the Social and Behavioral Sciences Institutional Review Board at the University of Chicago. Participants were not individually compensated for their participation; rather, they had the opportunity to enter (via email) one of several \$50 raffles (one drawing per 50 participants).

Participants were not actively recruited by the experimenters to complete the AP assessment (as noted above, most subjects participated following a mention of the test in the Wall Street Journal); thus, we did not employ traditional stopping rules for data collection. Rather, our sample size was based primarily on the length of time the AP assessment had been available online. That being said, it should be noted that the number of analyzable participants ($n = 152$) is approximately three times larger than a previous study providing evidence that AP is a continuously distributed ability [18]. However, it is smaller than one previous study arguing for AP as a dichotomous ability [7]. Moreover, based on an *a priori* power analysis using G*Power, the present sample size is sufficiently powered ($\beta = .8$) to detect medium effect sizes even when dividing the sample into three groups; i.e., genuine, intermediate, and non-AP ($f = 0.254$). The planned analysis that was used in the power analysis was a one-way ANOVA (e.g., assessing differences for age of music onset across groups); the power analysis did not specifically inform the Gaussian Mixture Modeling described in Section 3.4.

2.2 Materials

The experimental script was coded in jsPsych (de Leeuw, 2015). We tested absolute pitch-labeling ability with 48 complex tones: 24 “smooth” tones and 24 “triangle” tones. The smooth tones were generated using the “inverted sine” option in Adobe Audition (Adobe Systems: San Jose, CA), which did not result in a true sinusoid but rather a complex tone with 9 harmonics (in addition to the fundamental frequency) and an approximate 11 dB reduction for each harmonic. The triangle tones were also generated in Adobe Audition. Both the smooth tones and triangle tones spanned two octaves, from C [3] to B [4]. Additionally, both the smooth tones and triangle tones were 500ms in duration with a 50ms linear onset and offset ramp, and were RMS normalized to -5 dB full spectrum. A representative waveform and spectrum of both the smooth and triangle tone is presented in Fig 1.

2.3 Procedure

After providing informed consent, participants were presented with a 10-second sample of pink noise—normalized to the same level as the complex tones—and were instructed to adjust their computer’s volume to a comfortable listening level. After this volume adjustment, participants were instructed that they would hear an isolated note on each trial and must identify the name of the note within 5 seconds. After this instruction screen, participants completed the AP test. The 48 notes were presented in a pseudo-random order, as octave always switched between consecutive trials (though the selection of smooth or triangle timbre was random). The choice to continually switch octaves, meant to discourage relative pitch strategies, was influenced by prior research [16], including computerized tests of AP [18]. Participants responded by clicking on one of twelve buttons labeled with a musical note name on the screen

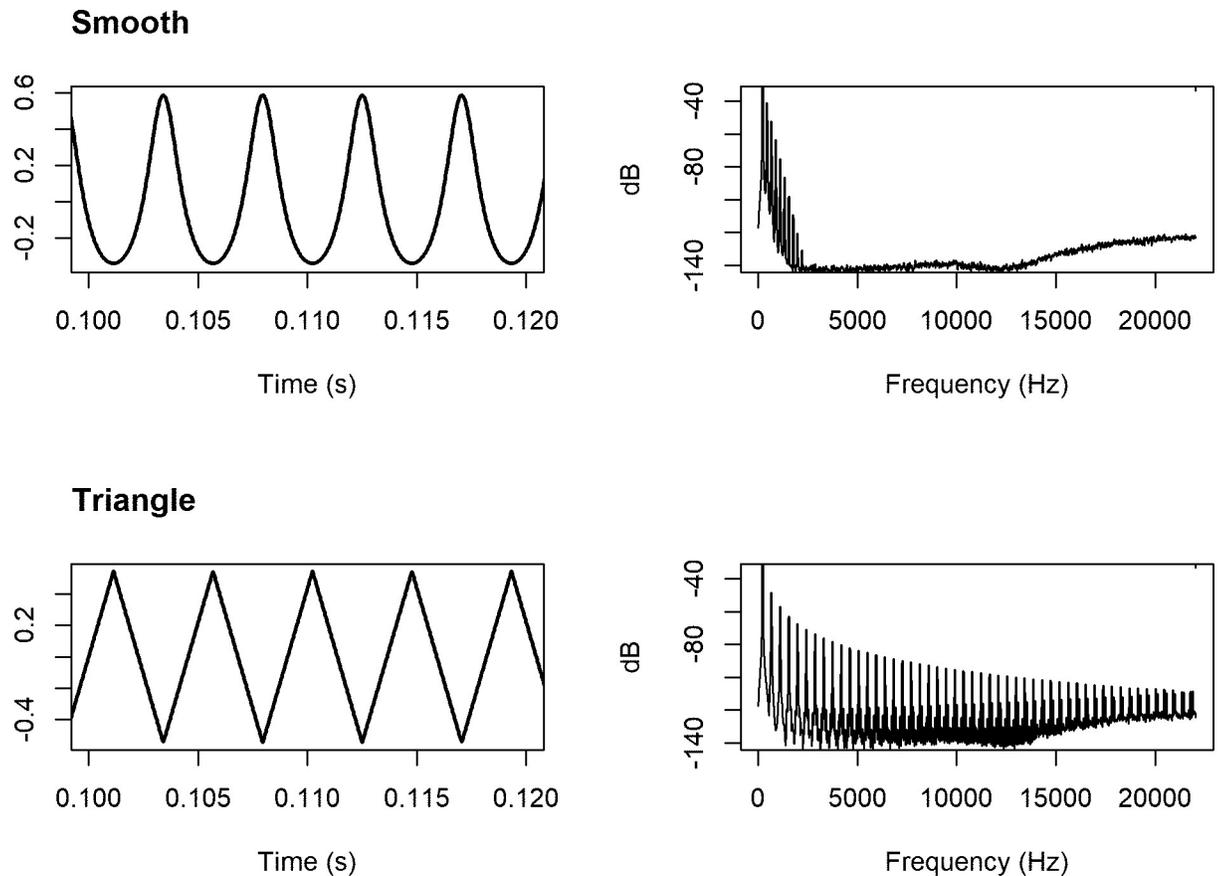


Fig 1. Waveform (left) and harmonic spectrum (right) for the smooth tone (top) and the triangle tone (bottom). Plots were generated from actual tones used in the AP assessment (A3; 220Hz).

<https://doi.org/10.1371/journal.pone.0244308.g001>

(e.g., F#). If participants did not respond within 5 second of the note, the trial was marked as a “time-out” and the experiment automatically advanced. As such, once participants advanced to the AP test, there was no opportunity to pause.

After the AP test, participants were presented with a brief questionnaire. Specifically, participants were asked to provide their current age, the age at which they began musical instruction (if applicable), and whether they spoke a tone language. Participants also had the opportunity to provide their email address to be entered into a monetary raffle for completing the study. However, these questions were non-obligatory, and participants could advance beyond this questionnaire page without entering any information. The final screen of the experiment provided participants with feedback (number of correctly identified notes, number of notes missed by one semitone, and percentage correct). The AP assessment took participants approximately five and a half minutes to complete, and most participants completed the entire procedure (from consent to the feedback screen) in under 10 minutes.

2.4 Dependent variables

The primary dependent measure we used was a composite score, incorporating both log response time (logRT) and mean absolute deviation (MAD) in semitones from the correct note into a single variable. Both logRT and MAD represent more graded measures of note categorization performance (as they extend beyond the “correct / incorrect” binary), and thus

both are important variables in understanding variability in AP performance. The composite score of logRT and MAD was calculated by adding 10 to an individual's MAD and then multiplying this value by their logRT. For example, if an individual's MAD was 1.5 semitones from the correct note and their average response time was 3500 milliseconds, their composite score would be 40.76 $[(1.5 + 10) * \log(3500)]$. This composite measure has been previously adopted in AP research [18] and is beneficial in that it imposes a penalty for slower responses.

To assess how our results changed with more restrictive operationalizations of AP, we also calculated a secondary measure. This secondary measure, hereafter referred to as the “conservative measure,” adopted the scoring principles outlined in prior influential AP studies [12,22] that have proposed a threshold for exceptional pitch-labeling ability (referred to by the authors as “AP1”). Moreover, this measure has been used previously to support a dichotomy in AP performance [7]. The conservative measure only considered responses made within a 3-second time window (as this was the reported note presentation rate from Baharloo et al. [12]), granting full credit for a correct answer, and granting three-quarters credit for answers that were removed by one semitone from the correct answer. Answers that were removed from the correct answer by more than one semitone were treated equivalently and not granted credit.

Trials in which participants failed to log a response within the 5-second window (i.e., “time-outs”) were marked as incorrect and were also assigned a conservative response time of 5 seconds. However, given that no response was logged for time-out trials, MAD could not be calculated.

2.5 Data analysis

Our first analysis goal was to characterize our data using the (often arbitrary) performance thresholds that are characteristic of AP research so that we could subsequently compare these groupings to a more data-driven categorization. To this end, we broke subjects into three groups based on their pitch classification accuracies. The break between chance-level, non-AP performers (nAP) and above-chance performers was set at 11 of 48 correctly identified notes (22.9%) because achieving this level of accuracy or higher significantly differed from the chance estimate of 4 of 48 notes (Binomial Test: $p = .0003$), even when using a false discovery rate alpha correction (FDR $q = .0007$).

The above-chance performers were further subdivided into pseudo-AP (pAP) performers, whom the AP literature would traditionally not consider performing strongly enough to have AP, and genuine-AP (gAP) performers. The threshold between these two groups was set at 39 of 48 correctly identified notes (81.3%)—a value that was based on prior research. For example, Deutsch et al. [23] used a similar threshold (85%) and operationalized performance both liberally (including semitone errors) and conservatively (only including correct classifications). Furthermore, Miyazaki [24] only tested self-identified AP possessors; however, in this study, the mean accuracy for complex tones (like those used in the present study) was 80.4%. As such, while the placement of any threshold is likely to be arbitrary (assuming a continuous distribution of performance), our selected threshold is grounded in prior research. Whether it is appropriate to separate these two groups at all is tested later in our analyses.

These pitch-labeling thresholds were also balanced with the composite score (incorporating speed of classification and MAD) to determine the final thresholds for the three participant groups. The first group, referred to as “genuine-AP” (gAP), required correctly identifying at least 39 of 48 notes (81.3%) *and a composite measure under 35*. The second group, referred to as “pseudo-AP” (pAP), required correctly identifying at least 11 of 48 notes (22.9%) *and a composite measure under 40*. The accuracy threshold for determining pAP group membership was set at 11 of 48 notes because achieving this level of accuracy or higher significantly differed

from the chance estimate of 4 of 48 notes (Binomial Test: $p = .0003$), even when using a false discovery rate alpha correction (FDR $q = .0007$). The third group, referred to as “non-AP” (nAP), encompassed the remaining participants, i.e., correctly identifying fewer than 11 of 48 notes and/or a composite measure of 40 or higher. The gAP and pAP groups were robustly above chance performance, represented by 1/12 (8.33%). The nAP group, in contrast, did not differ from chance. Summary statistics (mean and range) are provided for each group in [Table 1](#).

It is common in the field to compare performance on white vs. black key notes as a measure of AP test sensitivity [18], since decades of research have demonstrated that AP performance is worse on black-key notes compared to white-key notes [16,25,26]. To test whether such a “white-key advantage” was present in the current study, we constructed a 2 (note: black, white) \times 3 (group: gAP, pAP, nAP) ANOVA, after which we conducted the appropriate post-hoc tests. This analysis was conducted using both accuracy and the composite measure.

Then, we assessed whether there were differences between these groups with respect to age of musical training onset and tonal language experience. First, we conducted a pairwise t -test between the nAP and gAP groups (for age of music onset) and a Fisher’s exact test (for tonal language experience) to assess whether we replicate the findings of past AP research when our data are dichotomized in the same way. Then, to assess this relationship for all three groups, we constructed two regression models—one with age of music onset as a continuous response variable and one with tonal language as a binary response variable—using group membership as a regressor. Since some theoretically important inferences could be drawn if the null hypothesis were true, but frequentist analyses do not allow inferences to be validly drawn about evidence for the null hypothesis, we subsequently took a Bayesian approach to complement our frequentist models. To determine the relative evidence in favor of the null versus the alternative hypothesis given the data, we calculated Bayesian equivalents of post-hoc comparisons among the three groups using JASP 0.9.0.1 (JASP Team, 2018; [27]). In the context of the present analyses, the reported BF (BF01) represents the relative evidence in favor of the *null* hypothesis (i.e., the compared groups do not differ). For example, a BF of 3 would mean that the observed data are 3 times more likely to occur under the null hypothesis, whereas a BF of 1/3 would mean that the data are 3 times more likely to occur under the alternative hypothesis. We conduct these analyses first using the composite measure, and then using the more conservative measure described above to ensure our results are robust to the operationalization of AP.

Finally, we characterized the distribution of pitch-labeling performance using data-driven groupings, and we contrasted these groupings with the arbitrary, literature-based groups we had used up until this point. To arrive at these empirical groupings, we used Gaussian Mixture Models (GMMs), implemented with the “mixtools” package in R [28]. By modelling the data as if it is drawn from k normal distributions, we can find the probability of each subject belonging to each group given the normal distributions that best explain the observed data (fit using an Expectation-Maximization algorithm). Importantly, the number of normal

Table 1. Summary statistics for the AP groups.

Group	Accuracy	MAD	RT (s)
gAP ($n = 42$)	93.6% [81.3% - 100%]	0.05 [0–0.39]	2.31 [1.60–3.25]
pAP ($n = 46$)	55.8% [22.9% - 83.3%]	0.48 [0.06–1.28]	3.16 [2.11–4.42]
nAP ($n = 64$)	9.3% [0% - 31.3%]	2.67 [0.71–3.57]	3.58 [2.35–4.48]

Note: Ranges of values are printed in brackets.

<https://doi.org/10.1371/journal.pone.0244308.t001>

distributions used to explain the data is not arbitrarily chosen by the researcher. Rather, this number is determined by a standard model selection criterion of minimizing the Bayesian Information Criterion (BIC). As such, we could assess whether the supposition of three AP groups is empirically supported across operationalizations of AP ability. Since the BIC supported the notion of two distributions, rather than three, we calculated the posterior probability that the predefined pseudo-AP group belonged in the high performing distribution with the genuine-AP performers (as opposed to the other distribution with the at-chance performers); in this way, we empirically test whether pseudo-AP represents a categorically different ability from genuine-AP or if these represent different gradations of the same ability. We again tested two operationalizations of AP performance—the composite measure and the conservative measure.

3. Results

3.1 AP performance

The distribution of performance, assessed by overall accuracy and the composite measure, is plotted in Fig 2A. “White-key advantage” was present in the current study, indicating that our test was sufficiently sensitive to assess AP performance. For overall accuracy (Fig 3A), we observed a significant main effect of note ($F(1, 149) = 33.78, p < .001, \eta_p^2 = 0.185$), with white-key notes being more accurately identified than black-key notes (white: 49.7%, black: 42.3%). There was also a significant note-by-group interaction ($F(2, 149) = 10.82, p < .001, \eta_p^2 = 0.106$), meaning the white-key advantage differed across AP groups. Post-hoc tests showed that this interaction was driven by the pAP group, which had a significantly larger white-key advantage (16.6%) than both the gAP (2.3%) and nAP (4.3%) groups (both p s $< .001$). The attenuated white-key advantage for the gAP group, however, may have been driven by near-ceiling accuracy. Under this explanation, we would expect white-key advantages to manifest in the composite measure, which consists of two non-binary measures and thus may be more sensitive to individual variation.

For the composite measure (Fig 3B), we similarly observed a significant main effect of note ($F(1, 149) = 13.63, p < .001, \eta_p^2 = 0.084$), with white-key notes having lower (better) composite scores than black-key notes (white: 38.88, black: 39.30). The composite measure also showed a significant note-by-group interaction ($F(2, 149) = 7.04, p = .001, \eta_p^2 = 0.086$), with post-hoc tests showing that the gAP (-0.78) and pAP (-0.93) groups had larger white-key advantages compared to the nAP (0.17) group ($p = .014$ and $p = .002$, respectively). These analyses, taken together, suggest that the gAP and pAP groups display worse performance on black key notes compared to white key notes, which is an expected pattern among “AP possessors” [25].

3.2 Relationship between early musical training, tone language, and AP performance

The importance of both critical-period musical training and tone language on AP is often supported by directly comparing “AP” and “non-AP” populations, finding that the AP individuals began musical instruction at an earlier age and are more likely speakers of a tone language compared to the non-AP individuals (e.g., see [4,29] for reviews). The present dataset replicates both findings using this dichotomized approach. The gAP group reported a significantly earlier age of music onset compared to the nAP group (5.76 years vs. 8.59 years, $t(99.1) = 4.74, p < .001$) and were more likely speakers of a tone language (12 of 42 participants vs. 8 of 64 participants, $p = .046$ Fisher’s Exact Test).

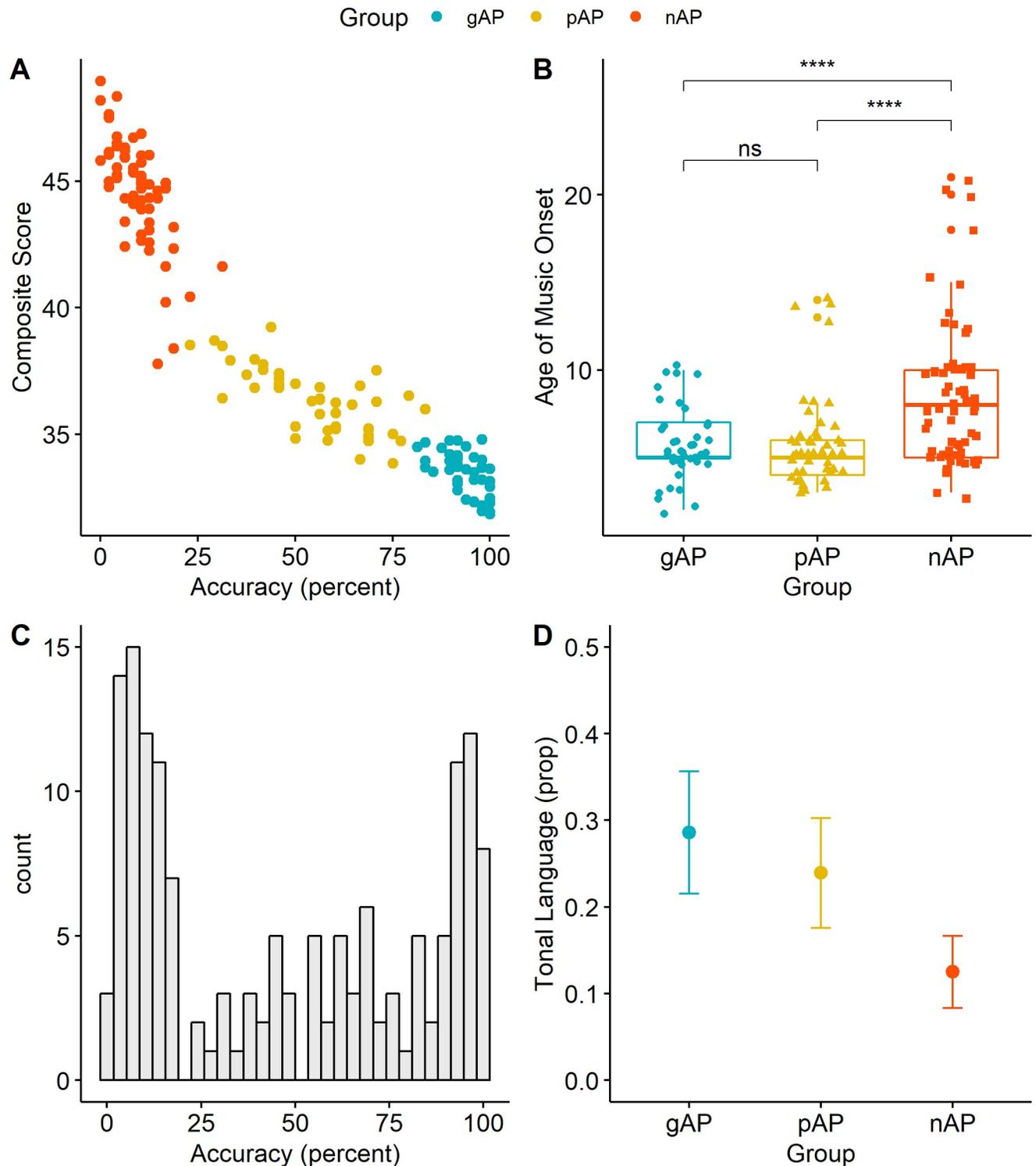


Fig 2. Scatterplot of AP performance across all participants, plotting mean accuracy (the percentage of trials in which a subject answered correctly) on the x-axis and the composite score incorporating MAD and logRT on the y-axis (A). The ages at which individuals in each group began musical instruction are shown, with quartiles for each group (genuine-AP, pseudo-AP, and non-AP), in (B). Distribution of pitch-labeling accuracies are shown in (C). Mean proportions of subjects that report speaking a tonal language are in (D), where error bars represent ± 1 standard error of the mean.

<https://doi.org/10.1371/journal.pone.0244308.g002>

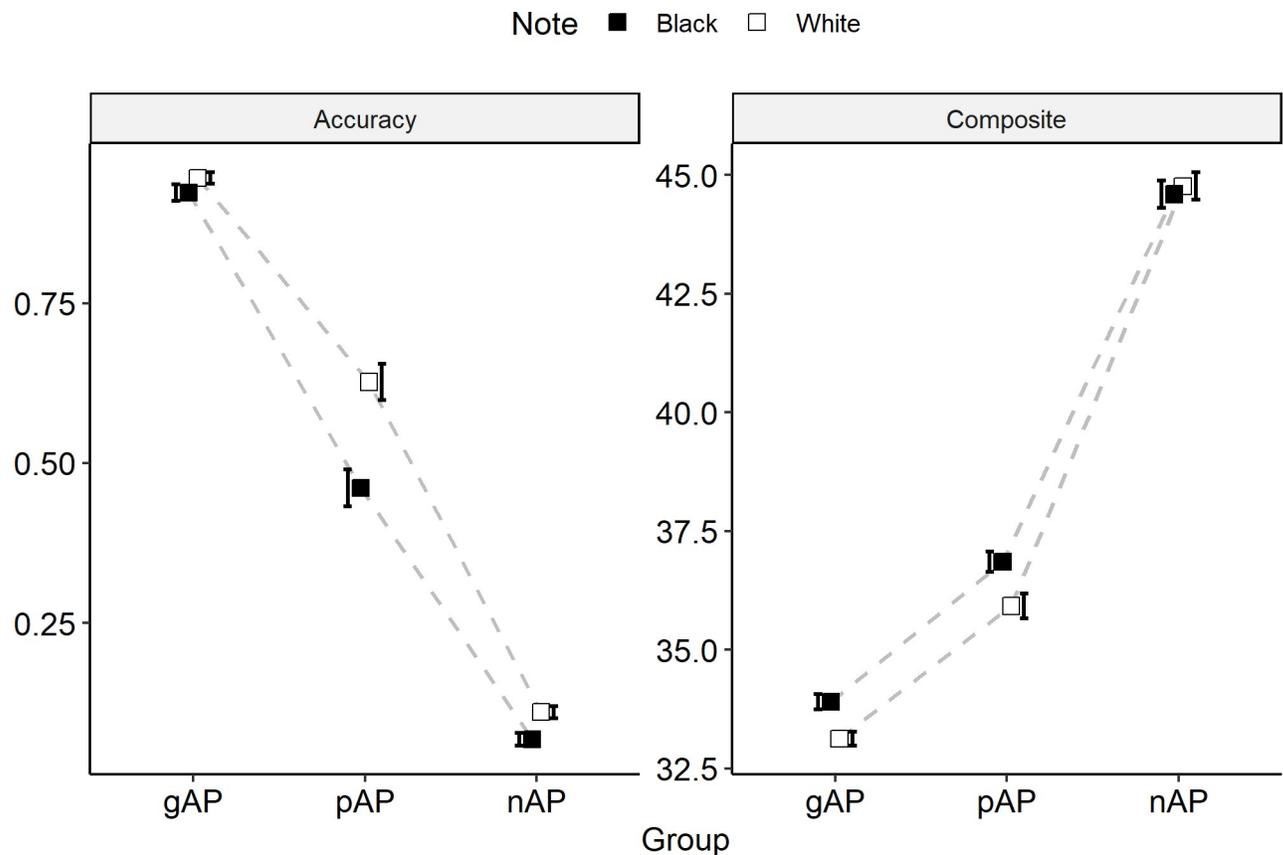


Fig 3. Interaction plots of from the 2 (note: White, black) x 3 (group: Genuine-AP, pseudo-AP, non-AP) ANOVAs for both mean accuracy and the composite measure. Error bars represent ± 1 standard error of the mean. Pseudo-AP subjects show the same performance bias known to be characteristic of genuine-AP level performers.

<https://doi.org/10.1371/journal.pone.0244308.g003>

The critical question for the present study, however, is how the pAP group relates to both the gAP and nAP groups in terms of critical-period musical training and tone language experience. The overall regression model for age of beginning musical instruction was significant ($F(2, 149) = 13.42, p < .001$). In terms of group differences, the pAP group reported a significantly earlier age of music onset compared to the nAP group ($B = 2.66, SE = 0.62, p < .001$) but did not significantly differ from the gAP group ($B = -0.17, SE = 0.69, p = .802$). Histograms of the age of music onset across groups are plotted in Fig 2B and mean values for each group are plotted in Fig 2C. The overall model for tone language was *not* significant ($F(2, 149) = 2.29, p = .105$). Despite the observation that the pAP group was almost twice as likely to speak a tone language compared to the nAP group (23.9% vs 12.5%), the difference was not significant ($B = -0.11, SE = 0.08, p = .143$). The difference between pAP and gAP participants was also not significant ($B = 0.05, SE = 0.09, p = .587$); however, the pAP and gAP groups were nominally more comparable (pAP: 23.9%, gAP: 28.6%). The mean proportion of tone language speakers across groups is plotted in Fig 2D.

These results suggest that the pAP group and the gAP groups had comparable ages of beginning musical instruction and proportions of tone language proportions, although these findings must be interpreted cautiously as they rest on the acceptance of a null hypothesis. To facilitate appropriate interpretation, since a null result here could be theoretically important but the regression above analyses provides us with little information about the posterior

probability the null hypothesis is correct, we computed Bayes Factors (BF01) to assess evidence in favor of the null.

For age of beginning musical instruction, the BF01 for the comparison of the pAP and gAP groups was 4.27, meaning the data were 4.27 times more likely to occur under the null hypothesis. In contrast, the BF01 for the comparison of the pAP and nAP groups was 0.008, meaning that the data were 125 times more likely to occur under the *alternative* hypothesis (i.e., that the pAP group and nAP group differed with respect to their mean age of beginning musical training). For tone language experience, the BF for the comparison of the pAP and gAP groups was 4.03, meaning the data were 4.03 times more likely to occur under the null hypothesis. Unlike the previous analysis, however, the BF01 for the comparison of the pAP and nAP groups was 1.64, meaning that the data were 1.64 times more likely to occur under the null hypothesis. Both of the BF01 values for the pAP/gAP contrasts can be interpreted as providing moderate evidence in favor of the null hypothesis [30], although the tone language results must be interpreted with caution as the BF did not support the alternative hypothesis when comparing the pAP and nAP groups.

3.3 Alternative operationalizations of AP performance

One potential concern with the present results is that we have chosen to operationalize AP in a specific manner (i.e., incorporating MAD and log response time into a single composite measure). While this measure has been used in the context of prior AP research [18], it is possible that this composite measure exaggerates the distributed nature of absolute pitch-labeling ability, as both components of the composite measure are non-binary. Given the variability in operationalizing AP across research groups, we thus assessed whether the interpretation of our results would change based on a different operationalization of AP performance.

As a strong test of our observation that AP is continuously distributed among above-chance performers, we chose an alternative operationalization of AP that has been previously used to argue for a dichotomy in AP performance [7]. The present study shares many similarities with the study reported by Athos and colleagues (including computerized testing and the use of non-musical timbres), giving face validity to such an approach. The details of how this conservative measure was calculated is reported in Section 2.4 (*Dependent Variables and Data Analyses*). Timbre is considered separately, meaning the maximum score is 24 (as each timbre consisted of 24 trials). Chance performance using this scoring measure is reflected by a score of 4.75, with the 95% confidence interval around chance performance encompassing the range [2.17, 7.89]. Using an identical equation as reported by the authors of this test to identify the highest (“genuine”) AP possessors, participants would need to score above 16.33 to be considered the highest (“AP1”) possessors.

We first assessed how our three participant groups—gAP, pAP, and nAP—scored using these performance criteria (Fig 5A). The first thing to note from Fig 4 is that performance on the inverted sine timbre was almost perfectly predictive of performance on the triangle timbre ($r(150) = .96, p < .001$). As such, we averaged inverted sine and triangle scores together for each participant to create a single score (out of 24). Not surprisingly, the gAP group had a high mean score ($M = 19.45, SD = 3.24$). However, seven of the 42 participants did not exceed the highest AP threshold of the test, highlighting its conservative nature relative to the composite measure. The nAP group scored essentially at chance ($M = 2.32, SD = 1.82$), with no participant exceeding the designated threshold for “AP1” and only one of 64 participants exceeding the upper limit of the 95% confidence interval around chance performance.

Even with this conservative measure, the pAP group scored between the gAP and nAP groups ($M = 10.69, SD = 4.66$)—i.e., they did not appear to be compressed toward the non-AP

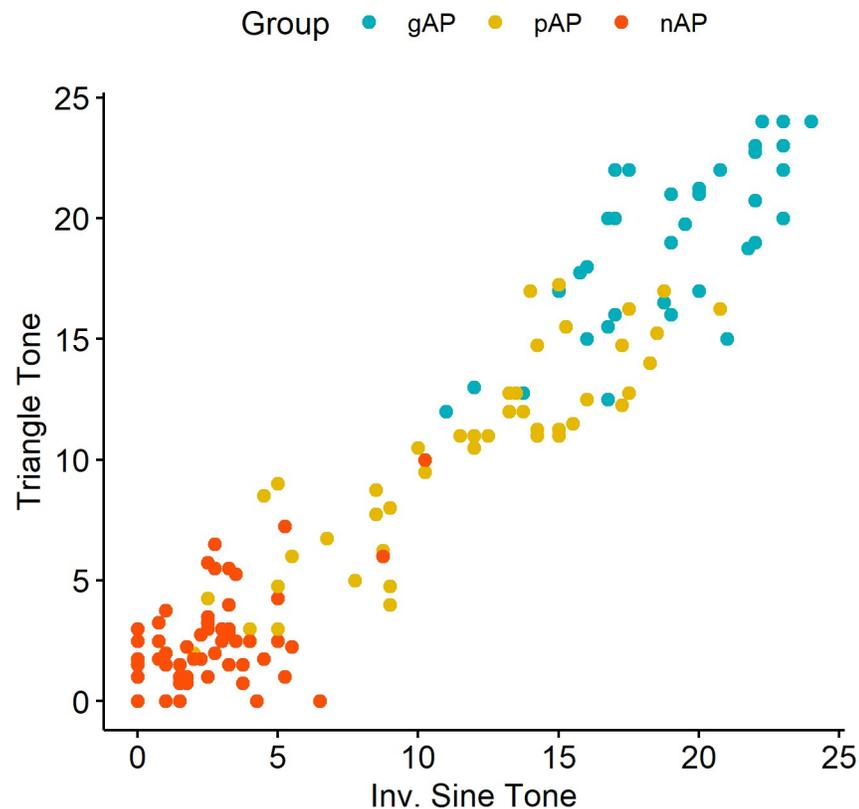


Fig 4. Scatterplot of performance on the inverted sine (x) and triangle (y) tones, using the conservative scoring measure. The three groups—genuine AP (gAP), pseudo AP (pAP), and non-AP (nAP) are carried over from the determination made using the composite measure.

<https://doi.org/10.1371/journal.pone.0244308.g004>

end of the distribution, as would be expected in a discrete framework. The pAP level of performance, on average, was almost perfectly between chance performance (5.94 points above) and the “AP1” threshold (5.64 points below). A majority (31 of 46) of the pAP participants exceeded the upper limit of the 95% confidence interval around chance performance, although only a small minority (4 of 42) exceeded the stringent “AP1” threshold. In sum, even using this conservative operationalization of AP, hypothesized to dichotomize performance based on previous work [7], we replicated our findings that AP performance is not discrete, as a sizable proportion of our sample (23.0%) performed at a level greater than expected by chance but below the “AP1” threshold. In fact, the proportion of participants performing at this intermediate level of AP was roughly comparable to the proportion of individuals who were classified as “AP1” (25.7%).

3.4 Modelling the distribution of AP performance across measures

While the gAP, pAP, and nAP groupings are useful for facilitating interpretation of our results considering previous research, they nonetheless represent a clear supposition about how AP performance is distributed. To mitigate the possibility of such arbitrary grouping influencing the interpretation of our results, we conducted further analyses using empirical, rather than *a priori*, groupings of the data. Since Bayesian Information Criteria supported two groups for both AP operationalizations tested, we were able to assess whether intermediate performers (pseudo-AP) were more naturally grouped with genuine-AP performers, suggesting that pseudo-AP and genuine-AP are just gradations of the same ability, or with the low performers.

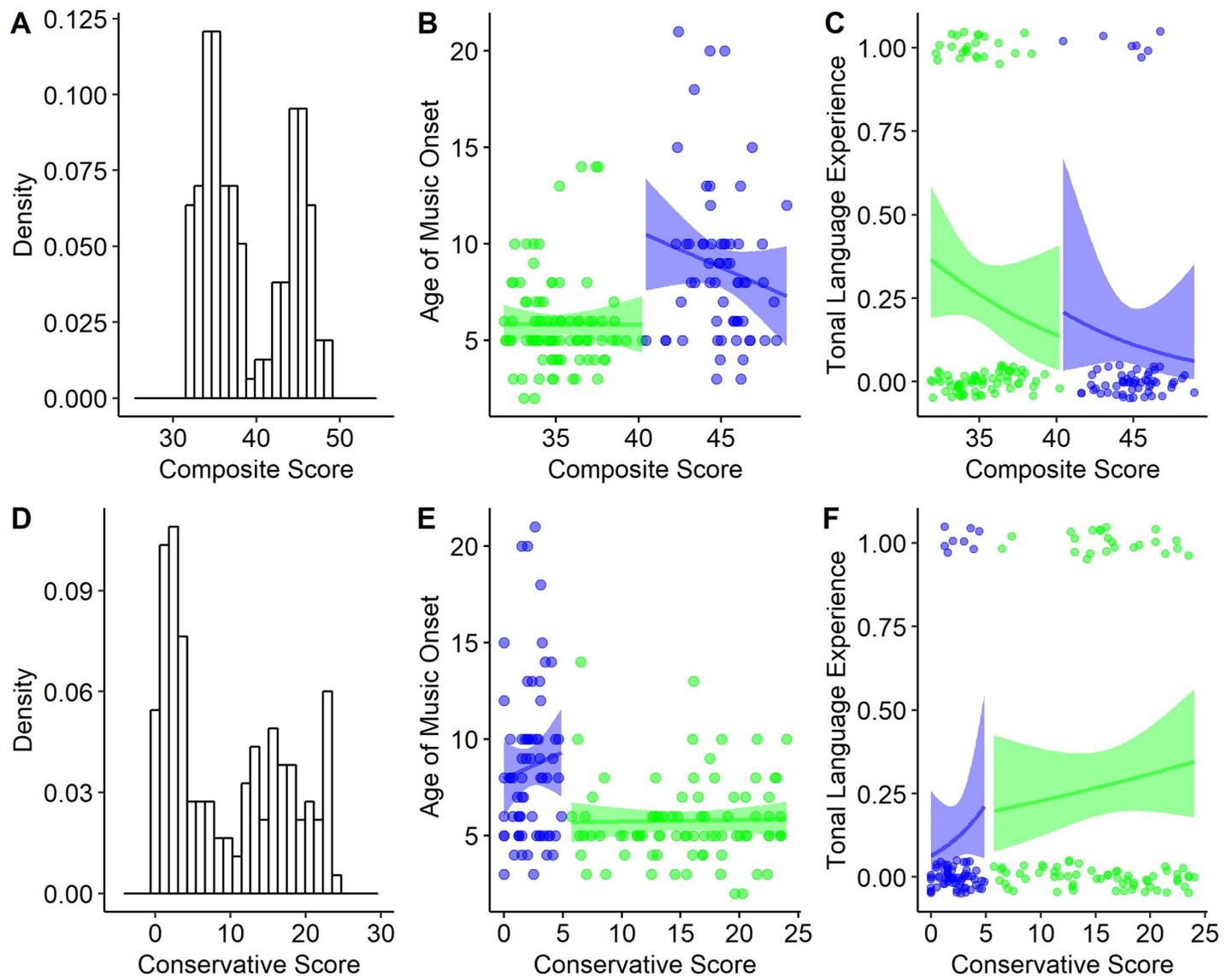


Fig 5. Results from the distributional modelling of AP performance using Gaussian Mixture Models (GMMs). 5A: GMM results from the composite score, depicting the best-fit distributions of the high-performing (green) and low-performing (blue) groups, overlaid on a histogram of the AP composite score. 5B: Age of musical training onset plotted against composite score; points are colored according to the distribution they are more likely to belong to (posterior probability > 0.5), and lines of best fit are shown for each distribution. The line is virtually flat for the high performing distribution, suggesting that age of beginning musical training does little to differentiate gradations of pitch-labeling ability, consistent with the results shown in Fig 2. 5C: Tonal language experience vs. composite score; logistic best fit, representing the probability of speaking a tonal language given the composite score, is shown for each distribution. 5D-F: The same plots as 5A-C but using the more conservative operationalization of AP instead of composite score.

<https://doi.org/10.1371/journal.pone.0244308.g005>

3.4.1 Composite score. For the composite score, the best-fitting model only included two groups. The increase in BIC from a two- to three-group model was 6.19. Applying a two-group GMM to the composite scores, we find that the distribution of the scores are best described as a mixture of the two normal distributions $N(35.06, 3.66)$ and $N(44.99, 3.04)$, shown in Fig 5A. The normal distribution with the lower mean composite score (representing superior performance) is taken to represent the AP group, and the distribution with the higher mean score is taken to represent the non-AP group. An advantage of GMMs is that they allow for overlapping clusters, but our model nonetheless makes group assignments with a high degree of

certainty, suggesting that the AP and non-AP distributions do not overlap substantially. Interestingly, these two distributions are well aligned with the mean composite scores of the AP (35.81) and non-AP (46.92) groups described in Bermudez and Zatorre [18].

Given that a two-group model, on the surface, may be thought to support a dichotomy in AP performance, the critical question is how the intermediate performers (i.e., the pAP group) is represented within these two distributions. To this end, a convenient summary statistic to use is the average probability of someone who is typically considered to have pseudo-AP performance coming from the AP group in our model. This average posterior probability of being part of the AP group rounds to 1.00 for the pAP participants (0.999), supporting the notion that pseudo-AP possessors are drawn from the same distribution as genuine AP possessors rather than the non-AP distribution. While this number should, by no means, be taken as the probability of the continuously distributed (among above-chance performers) hypotheses being true—indeed, there are many alternate hypotheses and distributional forms not considered here—it resoundingly favors the hypothesis that in two-group characterizations of AP, intermediate performance is better explained as part of the more accurate distribution of ability (including “genuine AP” performers), as opposed to the less accurate distribution of ability (encompassing at-chance, “non AP” performers).

3.4.2 Conservative measure. For the conservative measure, the best-fitting model was also a two-group model, which is perhaps not surprising as this measure has been previously used to support a dichotomy in AP. The three-group model provided a worse fit (increase in BIC of 8.22). Applying a two-group GMM to the conservative measure, we find that the distribution of the scores are best described as a mixture of normal distributions $N(2.18, 1.36)$ and $N(15.28, 5.50)$, shown in Fig 5B. These two distributions nominally can be thought of as a worse-performing, “non-AP” distribution, and a higher-performing, “AP” distribution. Similar to the composite measure analysis, we assessed the posterior probability of the pAP group being represented in the AP distribution. While the results were less decisive than the composite measure, the average posterior probability of belonging to the AP distribution for pAP participants was 0.8467, similarly supporting the notion that pseudo-AP possessors are more likely to belong to the same distribution as genuine AP possessors compared to non-AP possessors.

4. Discussion

Accurate descriptions of behavior are fundamental to developing explanations of the behavior. Thus, the way a behavior is distributed in the population is critical for informing any theory of underlying mechanisms and subsequent advances in characterizing the phenotype. When AP is conceptualized as simply present or absent, explaining AP requires theoretic mechanisms that are themselves rarely invoked and must combine in such a way as to materialize the ability without variability. Moreover, this discrete framework shapes the basic scientific questions in particular ways, limiting the range or characterization of participants who are included in research.

The present results directly challenge the hypothesis that AP is a categorically discrete ability, characterized by virtually no performance variability between clusters of traditionally defined “AP” and “non-AP” participants [7]. To highlight this point, 52 (30.6%) of all respondents performed with sufficient speed and accuracy to be distinguished from chance performance yet not considered to represent “genuine” AP ability based on commonly adopted thresholds (e.g., minimum of 85% accuracy). Simply discarding these participants and only comparing the extremes of the AP spectrum provided support for role of critical-period musical training and tone language experience in pitch-labeling, replicating prior research in terms

of potential mechanistic explanations of AP. Importantly, however, the present results clearly demonstrate that intermediate levels of AP look more like “genuine” AP in terms of the experiential signatures of critical-period musical training, tone language experience, and even inter-note performance differences such as white-key advantages. As such, these factors appear to relate to AP performance, but they derive their predictive value in differentiating “at-chance” from “above-chance” individuals, not the “genuine” AP possessors from all other individuals.

The stark dichotomy proposed in the case of AP—which is at least tacitly endorsed in AP research through the paradigmatic contrast of “AP” and “non-AP” groups as representative categories—has shaped the kinds of questions and results that have been reported. For example, studies examining the extent to which AP can be explicitly trained have generally supported the idea that AP performance can be improved in both children [31] and adults [32–35] to varying degrees. Yet, these demonstrable performance improvements are not seriously considered in the context of “genuine” AP ability—particularly among adults—given the discrete framing of AP in combination with critical period theories of AP development. In essence, these factors treat AP as a fixed or crystallized talent, and thus any meaningful improvements to pitch-labeling ability in adulthood must be explained through alternate mechanisms that mimic—but are distinct from—those underlying AP.

Moreover, this approach that has shaped much of the field is puzzling given the suggestions that there are likely multiple factors that contribute to the development of AP [36,37], which would suggest a more continuous distribution of ability. Furthermore, the finding that “AP possessors” tend to use multiple strategies in categorizing musical notes [38] supports the notion that, even among a conventionally-defined AP group, there may be significant variability in strategies and even overall ability. Understanding that AP varies within and between individual listeners is much more consistent with a cognitive process or skill [39] than to a kind of genetic endowment or a holistically acquired trait or talent.

While recent research supports the treatment of AP as a *continuously* distributed ability [17,18], intermediate AP abilities have received relatively little empirical attention in the broader context of AP research. Why might this be the case? One practical reason is that individuals with intermediate AP may not strongly self-identify as “AP possessors,” biasing recruitment efforts. One theoretical reason is that the discrete view of AP implicitly informs the design, analysis, and selection of participants for many AP studies. For example, intermediate levels of AP performance may be underrepresented due to how AP performance is tested and scored (e.g., treating all incorrect answers as equivalent, not modeling response time). Beyond biased recruitment and testing, even when intermediate AP is experimentally assessed, it is often assumed to represent a fundamentally different phenomenon than “genuine” AP [20,40]—ostensibly served by different mechanisms—rather than existing along the same AP continuum. By treating all individuals who do not reach this threshold as equivalent, it becomes difficult to understand the true distributional nature of AP, particularly if there is a substantial number of individuals who demonstrate intermediate AP performance across a variety of AP operationalizations (as in the present study).

Conceptualizing AP as a continuously distributed ability among above-chance performers may provide insights into understanding the nature of pitch memory in humans more broadly. In particular, a growing body of research has demonstrated that “non-AP” possessors display good absolute pitch memory for familiar sounds in their environment, such as popular music recordings [41–47] and even non-musical, pitched sounds, such as the North American land-line dial tone [48]. This kind of pitch memory, sometimes referred to as *implicit AP*, *latent AP*, or simply *pitch memory* [44,49,50], appears to be normally distributed in the population and has also been thought to represent a foundation for “genuine” AP (i.e., exceptionally accurate pitch memory may scaffold explicit pitch labeling). Supporting this idea, research has

demonstrated that AP possessors have more accurate “implicit” pitch memories compared to musically-matched controls [51], even when judging a non-musical stimulus in which their explicit note labels would not be beneficial [52]. Moreover, this pitch memory may reflect the effectiveness or precision of auditory working memory, given that auditory working memory appears to predict both the accuracy of “implicit” pitch memories for familiar recordings [47] (Van Hedger et al., 2017) as well as the explicit ability to learn AP categories [34].

Implicit pitch memory also appears to generalize beyond the specific recordings heard in one’s listening environment, which suggests a more abstracted representation of pitch chroma and provides a compelling connection to explicit AP. For example, non-AP listeners appear to respond differentially to isolated notes based on how frequently the notes are heard in the listening environment [53,54]. Relatedly, non-AP listeners can differentiate conventionally “in-tune” from “out-of-tune” notes with above-chance accuracy, though musicians outperform non-musicians and the effects are not observed for unfamiliar timbres [55]. These findings point to the intriguing possibility that the long-term representation of pitch chroma in memory generalizes beyond specific auditory experiences. The primary difference between pitch chroma representations between AP and non-AP listeners, then, may lie in the resolution of the representations, with “genuine” AP possessors displaying precise representations and non-AP possessors displaying broader representations (e.g., see [29]). It should be emphasized, however, that this kind of framework (differences in the resolution of pitch chroma) appears to be better represented as a continuum of ability.

Thus, the treatment of explicit AP labeling as an ability that is continuously distributed among above-chance performers opens an interesting line of potential research that may help to integrate implicit and explicit pitch memory research. Through the identification of intermediate AP performers, using tests such as the one administered in the present study, future work could for example test whether intermediate AP individuals display more accurate implicit pitch memories (e.g., differentiating in-tune and out-of-tune notes) compared to at-chance individuals, or whether implicit and explicit pitch memory abilities are independent among these populations. Such research approaches would promise to integrate explicit and implicit approaches to understanding how pitch is represented in long-term memory, as well as further clarify the distributional nature of AP.

Beyond illustrating that AP may be best described as an ability that is continuously distributed among above-chance performers, our results also challenge the view that intermediate levels of AP performance, often denigrated as “pseudo” AP, *necessarily* represents a fundamentally different process compared to “genuine” AP. In modeling the composite score, the most parsimonious representation of our data suggested by the Gaussian Mixture Model was one that only included two groups, and critically the intermediate performers were classified as belonging to the “genuine” AP distribution. Even when using a more restrictive and conservative scoring method, previously used to support a discrete view of pitch labeling ability [7,12,22], intermediate levels of pitch-labeling ability were observed in the data. Similar to the results from the composite score, these intermediate performers were much more likely to belong to the “genuine” AP distribution, not the “non-AP” distribution. In fact, the previously used threshold for AP from research groups using this operationalization arbitrarily bisected the AP distribution, highlighting the problem of using thresholds to differentiate “genuine” AP possessors from all other individuals.

In the present study, the use of uncommon timbres minimizes the possibility that the intermediate performers were only performing above chance because of extreme timbral familiarity, as can sometimes be the case for highly familiar timbres such as piano tones [56]. An additional concern with the use of complex timbres—as opposed to sine tones—in AP research is that *dynamic* spectro-temporal changes across pitch ranges (e.g., differences in the relative

power or temporal decay of harmonics) may provide an additional cue for successful note identification [57]. However, this is not applicable in the present study, as the amplitude envelope and relative power of the harmonics were fixed across all notes. Moreover, the choice to interleave octaves and provide no feedback during the assessment has been previously used to discourage the use of alternate, relative pitch strategies [18]. Furthermore, incorporating the speed of categorization into the pitch-labeling score was meant to penalize slower, deliberate response strategies that have been associated with relative pitch processing. As such, it is not likely that individuals were able to effectively use relative pitch to complete the task, at least based on the assumption that these design choices—rooted in previous research—represent valid discouragements of relative pitch strategies.

An important remaining question is thus—what factors explain performance variability among above-chance participants? One possibility is continued or specific musical expertise. Active musical training has been shown to improve pitch-production ability [58] and certain musical experiences, for example, playing a “variable *do*” instrument, can be detrimental to AP [37]. Unfortunately, given the limited nature of the questionnaire in the present study, we are unable to comment directly on how (continued) musical training relates to variability among the above-chance performers. A second possibility is (non-musical) auditory memory. For example, auditory working memory (WM) has been positively associated with the explicit learning of AP categories [34], with some high WM “pseudo” AP individuals demonstrating “genuine” AP levels of performance after eight weeks of training [35]. Furthermore, “genuine” AP possessors appear to have a larger auditory (but not visual) digit span compared to musically matched controls [59], though it is unclear how intermediate AP performers would compare to these conventional AP and non-AP groups. Third, it may be informative to use proposed biomarkers of AP-level pitch labeling ability in a more continuous framework. This approach has already displayed some promising results; for example, behavioral AP performance was found to relate significantly to the volume of white matter tracts connecting left superior temporal gyrus with left middle temporal gyrus [60].

The present results strongly argue against a fundamental *discrete dichotomy* of pitch-labeling ability; however, they do not inherently argue against a bimodal distribution of AP performance. Discrete performance implies that observations take on one of two values (e.g., “yes” or “no” with respect to possessing AP) which, given typical performance threshold for defining AP, suggests that virtually no observations should lie between these discrete groups. A more continuously distributed view of AP, in contrast, suggests that meaningful variability exists with respect to the ability to name an isolated musical note, but it also does *not* require that distribution to be uniform, unimodal, or normally distributed. While it could be argued that a bimodal distribution of AP performance still supports a dichotomy in AP, albeit loosely defined, the present data suggest that adopting such a stance would require a considerable expansion of what is typically thought to constitute AP (i.e., adopting more liberal performance thresholds that capture intermediate performance levels). Put another way, if one were to interpret the present results in a discrete framework, the data-driven approach of the present results suggest that *any* above-chance note naming performance should be included in AP investigations. In this sense, the present results fit within an emerging body of research that has demonstrated a bimodality of pitch-labeling ability, with considerable variability (i.e., a wider distribution) for individuals who perform above chance [61–63].

4.1 Conclusion

The present study offers important insights with respect to characterizing the distribution and proposed mechanisms of AP. Specifically, our findings suggest that “genuine” AP performers

and intermediate performers who have been characterized as having “pseudo-AP,” are differentiated neither by the age in which they began musical training nor whether they speak a tonal language, the experiential factors the most commonly are thought to predict absolute pitch-labeling ability. In contrast, a Gaussian mixture modelling analysis strongly suggested that genuine-AP and pseudo-AP are merely gradations of the same ability, rather than representing discrete abilities as previously argued. Importantly, while the onset of musical training (and possibly tonal language, though this is more ambiguous in our data), do differentiate subjects with at-chance performance from subjects with above-chance pitch labeling ability, these factors do not seem to have a statistical relationship with *gradations* of absolute pitch-labeling ability.

In other words, the conceptualization of AP as a discrete ability in which subject either have near-ceiling or at-chance performance is not supported by the data, and this theoretic assumption may have misled researchers seeking to characterize the interactions between genes and experience mediating this behavioral phenotype for the better part of a century. We recommend that future studies operationalize pitch-labeling ability among above-chance performers as a continuously measured ability, rather than as a discrete trait, and the full distribution of the ability should be samples, rather than just the best and worst performers. While it may be appropriate to consider at-chance performers separately, since our results suggest that they are likely drawn from a different distribution, there appears to be meaningful variance in the gradations of pitch-labeling ability that is yet to be explained. In general, we believe that these results be taken as a cautionary tale; individual differences in perception and behavior should be considered continuously until proven otherwise, and representative samples should be used, lest progress in cognitive science be hindered by analyses framed using nonexistent dichotomies.

Acknowledgments

The authors would like to thank Alex Huang for programming the AP assessment.

Author Contributions

Conceptualization: Shannon L. M. Heald, Howard C. Nusbaum.

Data curation: Stephen C. Van Hedger.

Formal analysis: Stephen C. Van Hedger, John Veillette.

Funding acquisition: Howard C. Nusbaum.

Investigation: Stephen C. Van Hedger, Shannon L. M. Heald, Howard C. Nusbaum.

Methodology: Stephen C. Van Hedger, Shannon L. M. Heald.

Project administration: Stephen C. Van Hedger, Howard C. Nusbaum.

Resources: Stephen C. Van Hedger.

Software: John Veillette.

Supervision: Howard C. Nusbaum.

Visualization: Stephen C. Van Hedger, John Veillette.

Writing – original draft: Stephen C. Van Hedger, Howard C. Nusbaum.

Writing – review & editing: Stephen C. Van Hedger, John Veillette, Shannon L. M. Heald, Howard C. Nusbaum.

References

1. Wagner S. L. et al., "Higher cortisol is associated with poorer executive functioning in preschool children: The role of parenting stress, parent coping and quality of daycare.," *Child Neuropsychol.*, vol. 22, no. 7, pp. 853–869, 2016, <https://doi.org/10.1080/09297049.2015.1080232> PMID: 26335047
2. Mankel K. and Bidelman G. M., "Inherent auditory skills rather than formal music training shape the neural encoding of speech," *Proc. Natl. Acad. Sci.*, vol. 115, no. 51, pp. 13129 LP– 13134, Dec. 2018, <https://doi.org/10.1073/pnas.1811793115> PMID: 30509989
3. Blagrove M., Hale S., Lockheart J., Carr M., Jones A., and Valli K., "Testing the Empathy Theory of Dreaming: The Relationships Between Dream Sharing and Trait and State Empathy," *Frontiers in Psychology*, vol. 10, p. 1351, 2019, [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsyg.2019.01351>. <https://doi.org/10.3389/fpsyg.2019.01351> PMID: 31281278
4. Deutsch D., *Absolute Pitch*. 2013.
5. Takeuchi A. H. and Hulse S. H., "Absolute pitch," *Psychol. Bull.*, vol. 113, no. 2, pp. 345–361, 1993, <https://doi.org/10.1037/0033-2909.113.2.345> PMID: 8451339
6. Zatorre R. J., "Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function.," *Nat. Neurosci.*, vol. 6, no. 7, pp. 692–695, 2003, <https://doi.org/10.1038/nn1085> PMID: 12830161
7. Athos E. A. et al., "Dichotomy and perceptual distortions in absolute pitch ability," *Proc. Natl. Acad. Sci.*, vol. 104, no. 37, pp. 14795–14800, 2007, <https://doi.org/10.1073/pnas.0703868104> PMID: 17724340
8. Bachem A., "The genesis of absolute pitch," *J. Acoust. Soc. Am.*, vol. 11, pp. 434–439, 1940.
9. Mull H. K., "The acquisition of absolute pitch," *Am. J. Psychol.*, vol. 36, no. 4, pp. 469–493, 1925, <https://doi.org/10.2307/1413906>
10. Sergeant D., "Experimental Investigation of Absolute Pitch," *J. Res. Music Educ.*, vol. 17, no. 1, p. 135, 1969, <https://doi.org/10.2307/3344200>
11. Levitin D. J. and Zatorre R. J., "On the Nature of Early Music Training and Absolute Pitch: A Reply to Brown, Sachs, Cammuso, and Folstein," *Music Percept.*, vol. 21, no. 1, pp. 105–110, 2003, <https://doi.org/10.1525/mp.2003.21.1.105>
12. Baharloo S., Johnston P. A., Service S. K., Gitschier J., and Freimer N. B., "Absolute Pitch: An Approach for Identification of Genetic and Nongenetic Components," *Am. J. Hum. Genet.*, vol. 62, no. 2, pp. 224–231, 1998, <https://doi.org/10.1086/301704> PMID: 9463312
13. Van Krevelin A., "The ability to make absolute judgments of pitch," *J. Exp. Psychol.*, vol. 42, no. 3, pp. 207–215, 1951, <https://doi.org/10.1037/h0062795> PMID: 14880674
14. Deutsch D., Henthorn T., and Dolson M., "Absolute Pitch, Speech, and Tone Language: Some Experiments and a Proposed Framework," *Music Percept.*, vol. 21, no. 3, pp. 339–356, 2004, <https://doi.org/10.1525/mp.2004.21.3.339>
15. Deutsch D., Dooley K., Henthorn T., and Head B., "Absolute pitch among students in an American music conservatory: Association with tone language fluency," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2398–2403, 2009, <https://doi.org/10.1121/1.3081389> PMID: 19354413
16. Deutsch D., Li X., and Shen J., "Absolute pitch among students at the Shanghai Conservatory of Music: A large-scale direct-test study," *J. Acoust. Soc. Am.*, vol. 134, no. 5, pp. 3853–3859, 2013, <https://doi.org/10.1121/1.4824450> PMID: 24180794
17. Van Hedger S. C. and Nusbaum H. C., "Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background," *Acta Psychol. (Amst.)*, vol. 191, no. October, pp. 251–260, 2018, <https://doi.org/10.1016/j.actpsy.2018.10.007> PMID: 30347313
18. Bermudez P. and Zatorre R. J., "A distribution of absolute pitch ability as revealed by computerized testing," *Music Percept. An Interdiscip. J.*, vol. 27, no. 2, pp. 89–101, 2009, <https://doi.org/10.1525/MP.2009.27.2.89>
19. Vitouch O., "Absolutist models of absolute pitch are absolutely misleading," *Music Percept. An Interdiscip. J.*, vol. 21, no. 1, pp. 111–117, 2003.
20. Bachem A., "Various Types of Absolute Pitch," *J. Acoust. Soc. Am.*, vol. 9, no. 2, pp. 146–151, 1937, <https://doi.org/10.1121/1.1915919>
21. Mitchell H., "Can perfect pitch be learned?," *Wall Street Journal*, New York, 2017.
22. Baharloo S., Service S. K., Risch N., Gitschier J., and Freimer N. B., "Familial aggregation of absolute pitch.," *Am. J. Hum. Genet.*, vol. 67, no. 3, pp. 755–758, 2000, <https://doi.org/10.1086/303057> PMID: 10924408

23. Deutsch D., Henthorn T., Marvin E., and Xu H., "Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period," *J. Acoust. Soc. Am.*, vol. 119, no. 2, p. 719, 2006, <https://doi.org/10.1121/1.2151799> PMID: 16521731
24. Miyazaki K., "Absolute Pitch Identification: Effects of Timbre and Pitch Region," *Music Percept. An Interdiscip. J.*, vol. 7, no. 1, pp. 1–14, 1989, <https://doi.org/10.2307/40285445>
25. Miyazaki K., "The Speed of Musical Pitch Identification Possessors by Absolute-Pitch Possessors," *Music Percept.*, vol. 8, no. 2, pp. 177–188, 1990.
26. Takeuchi A. H. and Hulse S. H., "Absolute-Pitch Judgments of Black- and White-Key Pitches," *Music Percept. An Interdiscip. J.*, vol. 9, no. 1, pp. 27–46, 1991.
27. Marsman M. and Wagenmakers E. J., "Bayesian benefits with JASP," *Eur. J. Dev. Psychol.*, vol. 14, no. 5, pp. 545–555, 2017, <https://doi.org/10.1080/17405629.2016.1259614>
28. Benaglia T., Chauveau D., Hunter D. R., and Young D. S., "mixtools: An R Package for Analyzing Mixture Models," *J. Stat. Software; Vol 1, Issue 6*, Oct. 2009, [Online]. Available: <https://www.jstatsoft.org/v032/i06>.
29. Levitin D. J. and Rogers S. E., "Absolute pitch: Perception, coding, and controversies," *Trends Cogn. Sci.*, vol. 9, no. 1, pp. 26–33, 2005, <https://doi.org/10.1016/j.tics.2004.11.007> PMID: 15639438
30. Jarosz A. F. and Wiley J., "What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors," *J. Probl. Solving*, vol. 7, no. 1, pp. 2–9, 2014, <https://doi.org/10.7771/1932-6246.1167>
31. Miyazaki K. and Ogawa Y., "Learning Absolute Pitch by Children," *Music Percept.*, vol. 24, no. 1, pp. 63–78, 2006, <https://doi.org/10.1525/mp.2006.24.1.63>
32. Brady P. T., "Fixed-Scale Mechanism of Absolute Pitch," *J. Acoust. Soc. Am.*, vol. 48, no. 4B, pp. 883–887, 1970, <https://doi.org/10.1121/1.1912227> PMID: 5480385
33. Cuddy L. L., "Practice Effects in the Absolute Judgment of Pitch," *J. Acoust. Soc. Am.*, vol. 43, no. 5, pp. 1069–1076, 1968, <https://doi.org/10.1121/1.1910941> PMID: 5648097
34. Van Hedger S. C., Heald S. L. M., Koch R., and Nusbaum H. C., "Auditory working memory predicts individual differences in absolute pitch learning," *Cognition*, vol. 140, pp. 95–110, 2015, <https://doi.org/10.1016/j.cognition.2015.03.012> PMID: 25909580
35. Van Hedger S. C., Heald S. L. M., and Nusbaum H. C., "Absolute pitch can be learned by some adults," *PLoS One*, vol. 14, no. 9, p. e0223047, Sep. 2019, [Online]. Available: <https://doi.org/10.1371/journal.pone.0223047> PMID: 31550277
36. Theusch E., Basu A., and Gitschier J., "Genome-wide Study of Families with Absolute Pitch Reveals Linkage to 8q24.21 and Locus Heterogeneity," *Am. J. Hum. Genet.*, vol. 85, no. 1, pp. 112–119, 2009, <https://doi.org/10.1016/j.ajhg.2009.06.010> PMID: 19576568
37. Wilson S. J., Lusher D., Martin C. L., Rayner G., and McLachlan N., "Intersecting factors lead to absolute pitch acquisition that is maintained in a 'Fixed do' Environment," *Music Percept.*, vol. 29, no. 3, pp. 285–296, 2012.
38. Zatorre R. J. and Beckett C., "Multiple coding strategies in the retention of musical tones by possessors of absolute pitch," *Mem. Cognit.*, vol. 17, no. 5, pp. 582–589, 1989, <https://doi.org/10.3758/bf03197081> PMID: 2796743
39. Heald S. L. M., Van Hedger S. C., and Nusbaum H. C., *Understanding Sound: Auditory Skill Acquisition*, vol. 232. 2017.
40. Bachem A., "Absolute pitch," *J. Acoust. Soc. Am.*, vol. 27, no. 6, pp. 1180–1185, 1955.
41. Frieler K., Fischinger T., Schlemmer K., Lothwesen K., Jakubowski K., and Müllensiefen D., "Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs," *Music. Sci.*, vol. 17, no. 3, pp. 334–349, 2013, <https://doi.org/10.1177/1029864913493802>
42. Jakubowski K., Müllensiefen D., and Stewart L., "A developmental study of latent absolute pitch memory," *Q. J. Exp. Psychol.*, vol. 70, no. 3, pp. 434–443, 2017, <https://doi.org/10.1080/17470218.2015.1131726> PMID: 26978128
43. Levitin D. J., "Absolute memory for musical pitch: Evidence from the production of learned melodies," *Percept. Psychophys.*, vol. 56, no. 4, pp. 414–423, 1994, <https://doi.org/10.3758/bf03206733> PMID: 7984397
44. Schellenberg E. G. and Trehub S. E., "Good pitch memory is widespread," *Psychol. Sci.*, vol. 14, no. 3, pp. 262–266, 2003, <https://doi.org/10.1111/1467-9280.03432> PMID: 12741751
45. Schellenberg E. G. and Trehub S. E., "Is there an asian advantage for pitch memory?," *Music Percept. An Interdiscip. J.*, vol. 25, no. 3, pp. 241–252, 2008.
46. Terhardt E. and Seewann M., "Aural Key Identification and Its Relationship to Absolute Pitch," *Music Percept. An Interdiscip. J.*, vol. 1, no. 1, pp. 63–83, 1983, <https://doi.org/10.2307/40285250>

47. Van Hedger S. C., Heald S. L. M., and Nusbaum H. C., "Long-term pitch memory for music recordings is related to auditory working memory precision," *Q. J. Exp. Psychol.*, pp. 1–13, 2017, <https://doi.org/10.1080/17470218.2017.1307427> PMID: 28856955
48. Smith N. A. and Schmuckler M. A., "Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. EL77–EL84, 2008, <https://doi.org/10.1121/1.2896106> PMID: 18396925
49. Bartlette C., Henry M. L., and Moore J., "Interaction of Relative Pitch Memory and Latent Absolute Pitch for Songs in an Ordered List," *Psychomusicology*, vol. 24, no. 4, pp. 279–290, 2015, <https://doi.org/10.1037/pmu0000064>
50. Deutsch D., "Absolute Pitch," in *The Psychology of Music*, 3rd ed., Deutsch D., Ed. San Diego, CA: Academic Press, 2013, pp. 141–182.
51. Dooley K. D., "Absolute Pitch and Related Abilities," *ProQuest Diss. Theses*, p. 139, 2011, [Online]. Available: http://proxy.bc.edu/login?url=http://search.proquest.com/docview/887708705?accountid=9673%5Cnhttp://bc-primo.hosted.exlibrisgroup.com/openurl/BCL/services_page??url_ver=Z39.88-2004&ft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+thes.
52. Van Hedger S. C., Heald S. L. M., and Nusbaum H. C., "What the [bleep]? Enhanced absolute pitch memory for a 1000 Hz sine tone," *Cognition*, vol. 154, pp. 139–150, 2016, <https://doi.org/10.1016/j.cognition.2016.06.001> PMID: 27289485
53. Ben-Haim M. S., Eitan Z., and Chajut E., "Pitch memory and exposure effects," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 40, no. 1, pp. 24–32, 2014, <https://doi.org/10.1037/a0033583> PMID: 23875573
54. Eitan Z., Ben-Haim M. S., and Margulis E. H., "Implicit Absolute Pitch Representation Affects Basic Tonal Perception," *Music Percept. An Interdiscip. J.*, vol. 34, no. 5, pp. 569–584, 2017, <https://doi.org/10.1525/mp.2017.34.5.569>
55. Van Hedger S. C., Heald S. L. M., Huang A., Rutstein B., and Nusbaum H. C., "Telling in-tune from out-of-tune: widespread evidence for implicit absolute intonation," *Psychon. Bull. Rev.*, vol. 24, no. 2, pp. 481–488, 2017, <https://doi.org/10.3758/s13423-016-1099-1> PMID: 27383616
56. Ward W. D. and Burns E. M., "Absolute pitch," in *The Psychology of Music*, 1st ed., Deutsch D., Ed. San Diego, CA: Academic Press, 1982, pp. 431–451.
57. Lockhead G. and Byrd R., "Practically perfect pitch," *J. Acoust. Soc. Am.*, vol. 70, no. 2, pp. 387–389, 1981, <https://doi.org/10.1121/1.386773>
58. Dohn A., Garza-Villarreal E. A., Ribe L. R., and Wallentin M., "Musical activity tunes up absolute pitch ability," *Music Percept.*, vol. 31, no. 4, pp. 359–371, 2014, <https://doi.org/10.1525/MP.2014.31.4.359>
59. Deutsch D. and Dooley K., "Absolute pitch is associated with a large auditory digit span: A clue to its genesis," *J. Acoust. Soc. Am.*, vol. 133, no. 4, pp. 1859–1861, 2013, <https://doi.org/10.1121/1.4792217> PMID: 23556554
60. Loui P., Li H. C., Hohmann A., and Schlaug G., "Enhanced cortical connectivity in absolute pitch musicians: A model for local hyperconnectivity," *J. Cogn. Neurosci.*, vol. 23, no. 4, pp. 1015–1026, 2011, <https://doi.org/10.1162/jocn.2010.21500> PMID: 20515408
61. Brauchli C., Leipold S., and Jäncke L., "Univariate and multivariate analyses of functional networks in absolute pitch," *Neuroimage*, vol. 189, pp. 241–247, 2019, <https://doi.org/10.1016/j.neuroimage.2019.01.021> PMID: 30639332
62. Leipold S., Greber M., Sele S., and Jäncke L., "Neural patterns reveal single-trial information on absolute pitch and relative pitch perception," *Neuroimage*, vol. 200, pp. 132–141, 2019, <https://doi.org/10.1016/j.neuroimage.2019.06.030> PMID: 31238164
63. Wengenroth M., Blatow M., Heinecke A., Reinhardt J., Stippich C., and Hofmann E., "Increased Volume and Function of Right Auditory Cortex as a Marker for Absolute Pitch," *Cereb. Cortex*, pp. 1127–1137, 2014, <https://doi.org/10.1093/cercor/bhs391> PMID: 23302811